

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 19:14:53 ; Search time 185 Seconds
(without alignments)
3844.173 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTOLVAQWTRATEPRYP.....WDEPKPILLCQYETLSDSE 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A Geneseq_29Jan04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	13178.5	99.7	2518	3 AAB40574	Ab40574 Human ORF
2	12978	98.2	2507	6 ABU61812	Abu61812 Human nuc
3	10796.5	81.7	2462	6 ABU61813	Abu61813 Mouse nuc
4	7866	59.5	1495	2 AAW18226	Aaw18226 Transcrip
5	4611	34.9	876	7 ADC35088	Adc35088 Human bre
6	4187	31.7	2453	3 AAB12454	Aab12454 HNRCR pro
7	4043	30.6	2343	3 AAB12453	Aab12453 Human HNR
8	1113	8.4	216	6 ABU70857	Abu70857 Human adi
9	1060	8.0	3502	4 ABB58382	Abb58382 Drosophil
10	960	7.3	184	6 ABU70980	Abu70980 Human adi
11	774.5	5.9	224	6 ABU70979	Abu70979 Human adi
12	718	5.4	619	2 AAR99738	Aar99738 Retinoid
13	569.5	4.3	3371	6 ABO07211	Abo07211 Human p53
14	569.5	4.3	3664	6 ABR47592	Abr47592 Breast ca
15	569.5	4.3	3664	6 ABO53027	Abo53027 Human put
16	565.5	4.3	3266	3 AAB42491	Aab42491 Human ORF
17	552.5	4.2	2406	6 ABB82733	Abb82733 Human KRC
18	550	4.2	1963	4 ABB62819	Abb62819 Drosophil
19	539.5	4.1	8991	6 ABU08487	Abu08487 S. pneumo
20	533	4.0	3080	4 ABB64877	Abb64877 Drosophil
21	523	4.0	2639	4 ABB51016	Abb51016 Novel hum
22	522.5	4.0	2536	4 AAU33195	Aau33195 Novel hum
23	519	3.9	3536	4 ABB65480	Abb65480 Drosophil
24	518	3.9	2665	4 AAM14533	Aam14533 Peptide #
25	518	3.9	2665	4 ABB33490	Abb33490 Peptide #

Db 901 SGRATTAKSGAPQDSDDSSATCSADEVDEREGDKNRLLSPRESLLTPTGDPANASPOK 960
QY 960 PLDLKQLKQRAAAIPIPIQVTKVHEPREDAAPTKPAPPPPPQNLQPEDSAPQPGSSP 1019
Db 961 PLDLKQLKQRAAAIPIQVTKVHEPREDAAPTKPAPPPPPQNLQPEDSAPQPGSSP 1020
QY 1020 RGSRSFAPPADKEAFAAAEAKLPGDPPCWTSGLPFPVPPREVIVKASHPADPSAFSYP 1079
Db 1021 RGSRSFAPPADKEAFAAAEAKLPGDPPCWTSGLPFPVPPREVIVKASHPADPSAFSYP 1080
QY 1080 PGHPLPLGLHDTARVLPRLPTTINSNPPPLISSAKHPSVLEROIKAISQGMVOLHVPYSE 1139
Db 1081 PGHPLPLGLHDTARVLPRLPTTINSNPPPLISSAKHPSVLEROIKAISQGMVOLHVPYSE 1140
QY 1140 HAKAPVGVPTMGLPLPMDPKKLAPFSGVKQEQLSPRGQAGPPESGLGVPYTAQASVLRGTA 1199
Db 1141 HAKAPVGVPTMGLPLPMDPKKLAPFSGVKQEQLSPRGQAGPPESGLGVPYTAQASVLRGTA 1200
QY 1200 LGSVPGGSIKGIPTSTRVPSDSAITVRGSI THGTADVLVYKGTITRIIGEDSPSRDLDRGR 1259
Db 1201 LGSVPGGSIKGIPTSTRVPSDSAITVRGSI THGTADVLVYKGTITRIIGEDSPSRDLDRGR 1260
QY 1260 EDLSLPGHVIYEGKKGHVLSYEGGMSVTQCSKEDGSSGPPHETAPKRTYDMMEGRVG 1319
Db 1261 EDLSLPGHVIYEGKKGHVLSYEGGMSVTQCSKEDGSSGPPHETAPKRTYDMMEGRVG 1320
QY 1320 RAISSASIEGLMGRALIPPERHSPHLLKEOHIRGSI THGTADVLVYKGTITRIIGEDSPSRDLDRGR 1379
Db 1321 RAISSASIEGLMGRALIPPERHSPHLLKEOHIRGSI THGTADVLVYKGTITRIIGEDSPSRDLDRGR 1380
QY 1380 REGTTPPPPPSRDLTAYKTQALGPLKLKPAHEGLVATVKEAGRSIHEIPREELRHTPEL 1439
Db 1381 REGTTPPPPPSRDLTAYKTQALGPLKLKPAHEGLVATVKEAGRSIHEIPREELRHTPEL 1440
QY 1440 PLAPRLKEGSI THGTPLKVDTCASITGSKKHVRSILGSPGRTFPVPHPLDMADARAL 1499
Db 1441 PLAPRLKEGSI THGTPLKVDTCASITGSKKHVRSILGSPGRTFPVPHPLDMADARAL 1500
QY 1500 ERACYEESLSRSGTASSGGSIARGAPVIVPELGPROSPLTYEDHGAPFAGHLPRGSP 1559
Db 1501 ERACYEESLSRSGTASSGGSIARGAPVIVPELGPROSPLTYEDHGAPFAGHLPRGSP 1560
QY 1560 VTMREPTPRLQEGSLSSKASQDRKLTSTPREIAKSPHSPTVPEHHPHIPSYEHLRLGVS 1619
Db 1561 VTMREPTPRLQEGSLSSKASQDRKLTSTPREIAKSPHSPTVPEHHPHIPSYEHLRLGVS 1620
QY 1620 GVDLYRSHIPLAPDPTSI PRGIPLDAAAAYILPRHLAPNPTYPHLYPPYPLIRGYPDTAAL 1679
Db 1621 GVDLYRSHIPLAPDPTSI PRGIPLDAAAAYILPRHLAPNPTYPHLYPPYPLIRGYPDTAAL 1680
QY 1680 ENRQTIINDYITTSQQMHNTATAMADRLMGLSPRESSLALNYAAGPRGIIDLSQVPH 1739
Db 1681 ENRQTIINDYITTSQQMHNTATAMADRLMGLSPRESSLALNYAAGPRGIIDLSQVPH 1740
QY 1740 LPVLVPTPTCTATAMDRLAYLPTAPQFSSRRSSPLSPGGGTHLTTKPTTSSSRERD 1799
Db 1741 LPVLVPTPTCTATAMDRLAYLPTAPQFSSRRSSPLSPGGGTHLTTKPTTSSSRERD 1800
QY 1800 RDRERDREREKSIITSTTTTVEHAPIWRPGTQSGSGSGSGSGSGSGSRSPASHSHAHQ 1859
Db 1801 RDRERDREREKSIITSTTTTVEHAPIWRPGTQSGSGSGSGSGSGSGSRSPASHSHAHQ 1860
QY 1860 HSPISPRTDALQORPSVLHNTCMKGIITAVEPSKPTVLRSTSTSPVRPAATFPATHC 1919
Db 1861 HSPISPRTDALQORPSVLHNTCMKGIITAVEPSKPTVLRSTSTSPVRPAATFPATHC 1920
QY 1920 PLGGTLGVDGYPTLMEPVLPKEAPRVARPERPADTGHAFLAKPPARSGLEPASSPSKGS 1979
Db 1921 PLGGTLGVDGYPTLMEPVLPKEAPRVARPERPADTGHAFLAKPPARSGLEPASSPSKGS 1980
QY 1980 EPRPLVPVSGHATIIARTPAKNLAPHASDPDPAPPASADPHREKTSQKPSIQLELR 2039

Db 1981 EPRPLVPVSGHATIIARTPAKNLAPHASDPDPAPPASADPHREKTSQKPSIQLELR 2040
QY 2040 SLGYHGSSYSPEGVEFVSPVSSPSLTHDKGLPKHLEELDKSHLEGELRPKQPGVVLKGE 2099
Db 2041 SLGYHGSSYSPEGVEFVSPVSSPSLTHDKGLPKHLEELDKSHLEGELRPKQPGVVLKGE 2100
QY 2100 AAHLPHLRPLPESQSSSPLLOTPAGVKGHQVVTTLAQHISEVITODYTRHHHQQLSAPL 2159
Db 2101 AAHLPHLRPLPESQSSSPLLOTPAGVKGHQVVTTLAQHISEVITODYTRHHHQQLSAPL 2160
QY 2160 PAPLSPFCASCPVLDLRRPPSDLYLPDPDHGAPARGSPHSEGGKSPENKTSVLGGGE 2219
Db 2161 PAPLSPFCASCPVLDLRRPPSDLYLPDPDHGAPARGSPHSEGGKSPENKTSVLGGGE 2220
QY 2220 DGIEPVSPPEGTEPGHRSRSAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFPSKLTESN 2279
Db 2221 DGIEPVSPPEGTEPGHRSRSAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFPSKLTESN 2280
QY 2280 SAMVSKKQEIINKLNTNRNEPEYNIISQPGTEIFNMPAITGTGLMTYRSQAVQEHASTN 2339
Db 2281 SAMVSKKQEIINKLNTNRNEPEYNIISQPGTEIFNMPAITGTGLMTYRSQAVQEHASTN 2340
QY 2340 MGLEAIIRKALMGKYDOWEESPLSANAFNPLNASASLPAAMPITAADGRSDHLLTSPGG 2399
Db 2341 MGLEAIIRKALMGKYDOWEESPLSANAFNPLNASASLPAAMPITAADGRSDHLLTSPGG 2400
QY 2400 GSKAKVSGRPSRKAAPGLASGDRPPSVSVSHSEGCNRRTPLTNRVWEDRPSAGS 2459
Db 2401 GSKAKVSGRPSRKAAPGLASGDRPPSVSVSHSEGCNRRTPLTNRVWEDRPSAGS 2460
QY 2460 TFPFYNPLIMRLQAGVMASSPPPGLPAGSGPLAGPHAWDEEPKLLCSQYETLSOSE 2517
Db 2461 TFPFYNPLIMRLQAGVMASSPPPGLPAGSGPLAGPHAWDEEPKLLCSQYETLSOSE 2518

RESULT 2
ABU61812
ID ABU61812 standard; protein; 2507 AA.
XX
AC ABU61812;
XX
DT 12-AUG-2003 (first entry)
XX
DE Human nuclear receptor corepressor SMRte.
XX
KW Human: SMRte; nuclear receptor corepressor; gene therapy; tissue typing;
KW cancer.
XX
OS Homo sapiens.
FH Key
FT Domain
FT Region
FT
FT /label= Amphipathic helix
FT /note= "residues 172, 179, 186, 194 and 200 make up the
FT hydrophobic heptad repeat"
FT
FT /label= SANT_A_domain
FT /label= SANT_B_domain
FT
XX US2003027137-A1.
XX
PD 06-FEB-2003.
XX
PF 27-MAR-2001; 2001US-00819104.
XX
PR 29-MAR-2000; 2000US-0193138P.
XX
PA (CHEN/) CHEN J D.
XX
PI Chen JD;

||||| 2374 SVSSVSEGCNRTPLTRVWEDRSSAGSTFPYPLIMRLQAGVMASPPPPGLAAGS 2433
 ||||| 2489 GPLAGPHHAWDEPKPLLCQSYETLSDSE 2517
 ||||| 2434 GPLAGPHHAWDEPKPLLCQSYETLSDSE 2462

RESULT 4
 AAW18226
 ID AAW18226 standard; protein; 1495 AA.
 XX AAW18226;
 AC AAW18226;
 XX 24-SEP-1997 (first entry)
 DE Transcriptional co-repressor SMRT.
 XX Silencing mediator for retinoic acid and thyroid hormone receptor; SMRT;
 KW transcriptional co-repressor.
 XX Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Region 1. .160
 FT /label= N-terminal region
 FT /note= "proline-rich domain"
 FT Region 773. .790
 FT /label= ERDR_region
 FT Region 812. .827
 FT /label= SG_region
 FT Region 1061. .1132
 FT /label= glutamine-rich region
 FT Region 1201. .1495
 FT /label= C-terminal_region
 FT Peptide 1330. .1376
 FT /note= "alternatively spliced insert not present in the
 original two-hybrid clone"
 FT
 XX WO9709418-A1.
 XX 13-MAR-1997.
 XX 24-JUL-1996; 96WO-US012371.
 XX 01-SEP-1995; 95US-00522726.
 XX (SALK) SALK INST BIOLOGICAL STUDIES.
 XX Evans RM, Chen JD;
 PI WPI; 1997-192894/17.
 DR
 XX New co-suppressor of steroid-thyroid hormone receptor activity - also
 PT methods for identifying compounds that relieve its suppressant effect
 PT and/or activate receptors.
 XX
 PS Claim 2; Page 40-45; 71pp; English.
 XX
 CC A novel receptor interacting factor (AAW18226) is designated SMRT, i.e.
 CC silencing mediator (co-repressor) for retinoic acid receptor (RAR) and
 CC thyroid hormone receptor (TR). Its association with RAR and TR both in
 CC solution and on DNA response elements is destabilised by ligand. The
 CC interaction of SMRT with mutant receptors correlates with the
 CC transcriptional silencing activities of receptors. In vivo, SMRT
 CC functions as a potent co-repressor. A GAL4 DNA binding domain fusion of
 CC SMRT behaves as a frank repressor of a GAL4-dependent reporter. These
 CC data identify a novel class of cofactor which is believed to represent an
 CC important mediator of hormone action. Full-length cDNA for SMRT has been
 CC isolated from a Hela library in a two-hybrid screen using a GAL4 DNA
 CC binding domain/RXR fusion protein as bait
 XX Sequence 1495 AA;
 SQ

Query Match 59.5%; Score 7866; DB 2; Length 1495;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1486; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1031 DKEAFAAEAOKLPDPCWTSGLPPVPREVIKASPHAPDPSAFYAPGCHPLPLGLHD 1090
 DB 9 DKEAFAAEAOKLPDPCWTSGLPPVPREVIKASPHAPDPSAFYAPGCHPLPLGLHD 68
 QY 1091 TARPVLPRPPTTISNPPPLISSAKHPSVLERQICAIISQGMVQLHVPVSEHAKAPVGPVTM 1150
 DB 69 TARPVLPRPPTTISNPPPLISSAKHPSVLERQICAIISQGMVQLHVPVSEHAKAPVGPVTM 128
 QY 1151 GLPLPMDPKLAPFSGVKOEQLSPRQAGPPESLGVPTAQEASVLRGTALGSPVGSITK 1210
 DB 129 GLPLPMDPKLAPFSGVKOEQLSPRQAGPPESLGVPTAQEASVLRGTALGSPVGSITK 188
 QY 1211 GIPSTRVPSDSAITYRGSITHTGTPADVLYKGIITRIIGEDSPSRLDRGREDSLPKGHVY 1270
 DB 189 GIPSTRVPSDSAITYRGSITHTGTPADVLYKGIITRIIGEDSPSRLDRGREDSLPKGHVY 248
 QY 1271 EGKKGHVLSYEGGMSVTQCKEDGRSSGPPHETAAPKRTYDMMEGRVGRASISASIEGL 1330
 DB 249 EGKKGHVLSYEGGMSVTQCKEDGRSSGPPHETAAPKRTYDMMEGRVGRASISASIEGL 308
 QY 1331 MGRAIPPERHSPHLLKEQHHRIGSITQGIIPRSYVEAQEDYLRREAKLLKREGTPPPPPPS 1390
 DB 309 MGRAIPPERHSPHLLKEQHHRIGSITQGIIPRSYVEAQEDYLRREAKLLKREGTPPPPPPS 368
 QY 1391 RDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPREELRHTPELPLARPLKEGS 1450
 DB 369 RDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPREELRHTPELPLARPLKEGS 428
 QY 1451 ITQGTPLKYDTGASTTGSKKHDVRSLSIGSPGRFTFPVHPVDVWADARALERACVYESLKS 1510
 DB 429 ITQGTPLKYDTGASTTGSKKHDVRSLSIGSPGRFTFPVHPVDVWADARALERACVYESLKS 488
 QY 1511 RGTASSSGSIARGAPVIVPELGKPRQSPLTYEDHGAPFAGHLPGRSPVTMEPTPRLQ 1570
 DB 489 RGTASSSGSIARGAPVIVPELGKPRQSPLTYEDHGAPFAGHLPGRSPVTMEPTPRLQ 548
 QY 1571 EGSLSKASQDRKLTSTPREIAKSPHSTVPEHHPHIIPYEHLLRGVSGVDLYRSHIPL 1630
 DB 549 EGSLSKASQDRKLTSTPREIAKSPHSTVPEHHPHIIPYEHLLRGVSGVDLYRSHIPL 608
 QY 1631 AFDPTSI PRGIPLDAAAAYLPRHLAPNTYPHLYPPYLRGYDPTAALENROTIIINDYI 1690
 DB 609 AFDPTSI PRGIPLDAAAAYLPRHLAPNTYPHLYPPYLRGYDPTAALENROTIIINDYI 668
 QY 1691 TSQOMHNTATAMAQRADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPTPGT 1750
 DB 669 TSQOMHNTATAMAQRADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPTPGT 728
 QY 1751 PATAMDRILAYLPTAPQPFSSRHSPLSPGGPHLTPTTTSSSERERDRDRDRDR 1810
 DB 729 PATAMDRILAYLPTAPQPFSSRHSPLSPGGPHLTPTTTSSSERERDRDRDRDR 788
 QY 1811 EKSILTSTTTVEHAPIWRPCTEQSSGSGSGSGSGSSSRPASHSHAHOSHSPISPTQDA 1870
 DB 789 EKSILTSTTTVEHAPIWRPCTEQSSGSGSGSGSGSGSSSRPASHSHAHOSHSPISPTQDA 848
 QY 1871 LQORPSVLHNTGMKGIIITAVEPSKPTVLARSTSTSSPVRPAATPPATHCPGLGTLDGVYP 1930
 DB 849 LQORPSVLHNTGMKGIIITAVEPSKPTVLARSTSTSSPVRPAATPPATHCPGLGTLDGVYP 908
 QY 1931 TLMPEVLLPKAPRVARPERPRADTGHAFIAPKAPARSGLEPASPSSKSGSEPRPLVPVSG 1990
 DB 909 TLMPEVLLPKAPRVARPERPRADTGHAFIAPKAPARSGLEPASPSSKSGSEPRPLVPVSG 968
 QY 1991 HATIARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFISOEELSLGVHSGSYSP 2050
 DB 969 HATIARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFISOEELSLGVHSGSYSP 1028

QY 2051 EGVEPVSPVSSPSLTHDKGLPKHLEBDEKSHLEGELRPKQPGVFKLGGEAAHLPHLRPLP 2110
Db 1029 EGVEPVSPVSSPSLTHDKGLPKHLEBDEKSHLEGELRPKQPGVFKLGGEAAHLPHLRPLP 1088
QY 2111 ESQPSPPLOTPAGVKGHORVVTLAQHISEVITQDYTRHHQQOISAPLPAPLYSPPGAS 2170
Db 1089 ESQPSPPLOTPAGVKGHORVVTLAQHISEVITQDYTRHHQQOISAPLPAPLYSPPGAS 1148
QY 2171 CPVLDDRPPSDLYLPPPDHGPARGSPHSEGGKSPENKTSVLGGGSDGIEPVSPREG 2230
Db 1149 CPVLDDRPPSDLYLPPPDHGPARGSPHSEGGKSPENKTSVLGGGSDGIEPVSPREG 1208
QY 2231 MTEPGHRSVAVPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESNSAMVSKKQEI 2290
Db 1209 MTEPGHRSVAVPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESNSAMVSKKQEI 1268
QY 2291 NKKLNTNHNREBEYNIISQGTIFNMPAITGTGLMTYRSQAVQEHASTNMGLEAIIRKAL 2350
Db 1269 NKKLNTNHNREBEYNIISQGTIFNMPAITGTGLMTYRSQAVQEHASTNMGLEAIIRKAL 1328
QY 2351 MGKYDOWEESPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGGGKAKVSGRPS 2410
Db 1329 MGKYDOWEESPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGGGKAKVSGRPS 1388
QY 2411 SRKAKSPAPGLASGRDPPSVSVHSEGCNRRTPLTNRVWEDRPPSAGSTPPFYNPLMR 2470
Db 1389 SRKAKSPAPGLASGRDPPSVSVHSEGCNRRTPLTNRVWEDRPPSAGSTPPFYNPLMR 1448
QY 2471 LQAGWASPPPLGSLPAGSGPLAGPHHAWDEEBKPLLCQOYETLSDSE 2517
Db 1449 LQAGWASPPPLGSLPAGSGPLAGPHHAWDEEBKPLLCQOYETLSDSE 1495

RESULT 5

ADC35088
ID ADC35088 standard; protein; 876 AA.

XX AC ADC35088;
XX AC ADC35088;
XX 18-DEC-2003 (first entry)
XX DE Human breast cancer antigen seq id 54.
XX KW breast cancer; breast cancer diagnosis; breast cancer antigen.
XX KW Homo sapiens.

XX PN US2003108888-A1.
XX PD 12-JUN-2003.
XX PF 15-MAY-2002; 2002US-00146473.
XX PR 15-MAY-2001; 2001US-0291150P.
XX XX

XX (LUDW-) LUDWIG INST CANCER RES.
XX PA Scanlan MJ, Gout I, Stockert E, Old LJ, Gure A, Chen Y;
XX DR WPI; 2003-829397/77.
XX DR N-PSDB; ADC35130.

XX PT Diagnosing breast cancer in subject by obtaining biological sample from
XX PT subject, contacting sample with breast cancer-associated polypeptides,
XX PT determining specific binding between polypeptides and agents in sample.
XX PS Example 2; SEQ ID NO 54; 173pp; English.

XX CC The invention describes a method of diagnosing breast cancer in subject
XX CC comprising contacting biological sample from subject with at least two
XX CC different breast cancer-associated polypeptides (I) encoded by nucleic
XX CC acid molecules (II) comprising sequence chosen from 42 fully defined
XX CC sequences as given in specification, determining specific binding between

CC (I) and agents in sample, where presence of the binding is diagnostic for
CC breast cancer. The method is useful for diagnosing breast cancer in a
CC subject. The sample is blood, lymph node fluid or breast discharge fluid.
CC This is the amino acid sequence of a breast cancer antigen.
XX SQ Sequence 876 AA;

Query Match 34.9%; Score 4611; DB 7; Length 876;
Best Local Similarity 94.7%; Pred. No. 7e-251;
Matches 873; Conservative 0; Mismatches 3; Indels 46; Gaps 1;

QY 1596 PHSVPEHHHPISPYEHLRLRGVSGVDLYRSHIPLAFDPTSPRGIPLDAAAYYLPRHL 1655
Db 1 PHSVPEHHHPISPYEHLRLRGVSGVDLYRSHIPLAFDPTSPRGIPLDAAAYYLPRHL 60
QY 1656 ANPYPHLYPYLIRGYPDYTAALNRQTIINDYITISQOMHNTATAMQADMLRGLSP 1715
Db 61 ANPYPHLYPYLIRGYPDYTAALNRQTIINDYITISQOMHNTATAMQADMLRGLSP 120
QY 1716 RESSLALNVAAGPRGIIDLQVPHLPVLVPPPTGPTATAMDLRLAYLPTAQPFSSRHSS 1775
Db 121 RESSLALNVAAGPRGIIDLQVPHLPVLVPPPTGPTATAMDLRLAYLPTAQPFSSRHSS 180
QY 1776 PLSPGGPHTLTKPTTTSSERDRDRDREREKSIILSTTTVEHAPIWRPGTEQSS 1835
Db 181 PLSPGGPHTLTKPTTTSSERDRDRDREREKSIILSTTTVEHAPIWRPGTEQSS 240
QY 1836 GSSGSGGGGSSSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKGITITAVEPSKP 1895
Db 241 GSSGSGGGGSSSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKGITITAVEPSKP 300
QY 1896 TVLRSTSTSSPVPRPAATFPPTHCPGLGGTLDGVYPTLMEPVLLPKEAPRVARPERPRA 1955
Db 301 TVLRSTSTSSPVPRPAATFPPTHCPGLGGTLDGVYPTLMEPVLLPKEAPRVARPERPRA 360
QY 1956 GHAFIAPKPPARGLEPASPSPSGSEPRPLVPVSGHATITARTPAKNLAPHASDPDPAPP 2015
Db 361 GHAFIAPKPPARGLEPASPSPSGSEPRPLVPVSGHATITARTPAKNLAPHASDPDPAPP 420
QY 2016 ASADPHREKTKSPFSIQEELRSILGVHGSYSPEGVEPVSPVSSPSLTHDKGLPKHLE 2075
Db 421 ASADPHREKTKSPFSIQEELRSILGVHGSYSPEGVEPVSPVSSPSLTHDKGLPKHLE 480
QY 2076 ELDKSHLEGELRPKQPGVFKLGGEAAHLPHLRPLPESQSSSPQLQTAPGVKGHORVVT 2135
Db 481 ELDKSHLEGELRPKQPGVFKLGGEAAHLPHLRPLPESQSSSPQLQTAPGVKGHORVVT 540
QY 2136 AOHSIVITQDYTRHHQQOISAPLPAPLYSPPGASCPVLDRRPPSDLYLPPPDHGPAPAR 2195
Db 541 AOHSIVITQDYTRHHQQOISAPLPAPLYSPPGASCPVLDRRPPSDLYLPPPDHGPAPAR 600
QY 2196 GSPHSEGGKRSPEPNKTSVLGGGSDGIEPVSPPEGMTPEGHRSRQAVYLLYRDGEQTEPS 2255
Db 601 GSPHSEGGKRSPEPNKTSVLGGGSDGIEPVSPPEGMTPEGHRSRQAVYLLYRDGEQTEPS 660
QY 2256 RMGSKSPGNTSQPPAFFSKLTESNSAMVSKKQEIINKLNTNHNREBEYNIISQGTIFN 2315
Db 661 RMGSKSPGNTSQPPAFFSKLTESNSAMVSKKQEIINKLNTNHNREBEYNIISQGTIFN 720
QY 2316 MPATITGTGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDOWEESPPLSANAFNPLNAS 2375
Db 721 MPATITGTGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDOWEESPPLSANAFNPLNAS 756
QY 2376 SLPAAMPITAADGRSDHTLTSPGGGGKAKVSGRPSRKRKAKSPAGLASGRDPPSVSVHS 2435
Db 757 -----GGGGKAKVSGRPSRKRKAKSPAGLASGRDPPSVSVHS 794
QY 2436 EGDGNRRTPLTNRVWEDRPPSAGSTPPFYNPLMRQAGWASPPPLGSLPAGSGPLAGPH 2495
Db 795 EGDGNRRTPLTNRVWEDRPPSAGSTPPFYNPLMRQAGWASPPPLGSLPAGSGPLAGPH 854
QY 2496 HAWDEEBKPLLCQOYETLSDSE 2517
|||

Db 855 HAWDEPKLLCSQYETLSDSE 876

RESULT 6
AAB12454
ID AAB12454 standard; protein; 2453 AA.
XX AC AAB12454;
DT 24-OCT-2000 (first entry)
XX DE HNRCR protein sequence.
XX KW Human; HNRCR; nuclear receptor coreceptor.
XX OS Unidentified.
XX PN CN1250094-A.
XX PD 12-APR-2000.
XX PF 06-OCT-1998; 98CN-00120919.
XX PR 06-OCT-1998; 98CN-00120919.
XX PA (XINH-) XINHANGPU FUDAN GENE ENG CO LTD SHANGHA.
XX PI Yu L, Tu Q, Zhao Y;
XX DR WPI; 2000-400830/35.
XX DR N-PSDB; AAA60630.
XX PT Preparation of new human kernon acceptor co-repressor coding series and the polypeptide.
XX PS Example 2; Fig 2; 58pp; Chinese.
XX CC The present invention describes a human homologue of nuclear receptor coreceptor (HNRCR). The present sequence represents an HNRCR protein sequence used in comparison with the human HNRCR
XX CC
XX CC
XX Sequence 2453 AA;

Query Match 31.7%; Score 4187; DB 3; Length 2453;
Best Local Similarity 41.0%; Pred. No. 2.1e-226;
Matches 1103; Conservative 348; Mismatches 796; Indels 444; Gaps 106;

QY 16 EPRYPHSLSYVQIARTHTDVLGLLEQ--HHGRDYASHLSGSIITQPQRRRPSLLSEFQ 73
DB 17 QSRYPHSHVQYTPFSARHQOEFAVPDYRSSHLEVSOASQLLOQQQQQQLRRRPSLLSEF 76
QY 74 PGNRSQELHLRPESHSLYPGLCKSEMEFTESKRPLELLDPDLLR-----PSPLLATGQ 128
DB 77 PGSDRQOE--RRSGYQFHGPGSPVDHDSLESKRPLEQVSDSHFORISAAVPLVHT-L 133
QY 129 PAG---SEDITKDRSLTGKLE-PVSPSPPHPTDPELEVPPLSLKEELIQNMDRVDREIT 184
DB 134 PEGLRSSANAKDPAGVCKHEAPSSPLSGQPCGDDQNAPSKLSKEELIQNMDRVDREIA 193
QY 185 MVEQQTISKUKKQOQLEEEAANKPEPEKVPSPPIESKHSLVQIYYDENRKAHAHRI 244
DB 194 KVEQQTILKUKKQOQLEEEAANKPEPEKVPSPPIESKHSLVQIYYDENRKAHAHRI 253
QY 245 LEGLGPOVELPLYNQPSDTQVHENIKINQAMKLLILYFKRNHARKQWKQKFCORYDQ 304
DB 254 FEGLGPKVELPLYNQPSDTQVHENIKTNQVMKLLILFKRNHARKQWKQKFCORYDQ 313
QY 305 LMEALEKKVRIENRRRAKESKREYEEKQPEIRKQRELOERMQSRVQGRGSLMS 364
DB 314 LMEAEKKVDRIENRRRAKESKREYEEKQPEIRKQRELOERMQSRVQGRGSLMS 372
QY 365 AARSEHEVSEIIDGLSEQNLENKOMQLAVIPMLYDADQOIRKFINMGLMADPMKVYK 424

Db 373 IARSEHEISEIIDGLSEQNENKQKQLSVIPPMFDEAORRVKFINMGLMEDPMKVYK 432
QY 425 DRQVMNMWSEQKETPREKFMQHPKNGFLIASFLERKTVAECVLYLYLTKNENYKSLVR 484
Db 433 DRQFMNVWTDHEKEIFKDKFIQHPKNGFLIASFLERKSVPCDVLVLYLTKNENYKALVR 492
QY 485 RSY-RRRGKSQQQQQQQQQQQQQQQQQQQQPMRPSQBEKDEKEKEKEAEBK--EEREPEVEND 541
Db 493 RNYKGRGRNQIARPSQBEKVEEK-----EEDKAETKEKEKEKDEKDEKDDKEDS 544
QY 542 KEDLLKEKTDGSDGDNDEKAVASKRKTANSQGRKGRITRSMANE--ANSEBAITP 598
Db 545 KETTKKORTAETAEPEEREQVTPRKTANSQGRKGRITRSMETSEAAAAAANAATE 604
QY 599 QOSAEI-----ASMELNESSRWTEEMETAKGLLEHGRNWSAIAARMVGSKTVSQCKNF 652
Db 605 EPPPLPPPPPEPISTEPVETSRWTEEMEVAKGLVEHGRNWAIAKMGVTKSEAOCKNF 664
QY 653 YFNYKGRQNLDELLOQHKLMEKERNARKKAPAAASEAAFPVPVVEEHEASVSG 712
Db 665 YFNYKGRHNLNLLQHKQKARKPREERDVSCVESVASTVSA---QEDEDIEAS--- 716
QY 713 NEEEMVEEAEALHASGNEVPRGCGPATVNNSSDTEIPSPH-TEAAK---DTGONGPK 768
Db 717 NEEENPEDSEG-----AENSDTESAPSPVPEAKKSSSENA-- 756
QY 769 PPATLGADGPPPGPPTPRRTSRAPTEPATGATGPTPPAPPSPSPAPPVVPKEKE 828
Db 757 -----ASRGNTPEVALEATDPAFCASP--SSAVPTTKPAERES 794
QY 829 EE-----TAAAPPVEGEEQ-----KPPAAELAVDTGKAEBPVKSECTEAE 871
Db 795 VEAQVTDASAETAEPMDVDHECGAEGSVLDPAPATK--ADSDVPEMOV-PENTASKG 851
QY 872 EGPAGKDAEAEATAEGALKAEKGGSGRATTAKSSGAPO--DSDSATSCADSDVE 928
Db 852 EGDAKERDLEST-----SEKTEARDEDVVVAQIERPEPQSDDDSATSCADSGVD 902
QY 929 AEGGDKNRL--LSRPSLLTPTGDRANASQPKP--LDLKLQKQRAAAIIPPI----- 976
Db 903 GE-PERQRFPMDAKPSLLTPGSIIL-SSPIKPNLLDLPLQHQRAAVIPPMVSCVPCNI 960
QY 977 -----QVTKVHEPPREDAATKAPPAPPAPPONLQPEDSAPQQCGSSPRGK 1022
Db 961 PIGTPVSGYALYORHIKAMHESALLE-----EQRQOEQVDLECRSSTSPCST 1008
QY 1023 SRSPAPDKEAFAAQAQKLPDPCWTSGLPPVPVPREVIKASPHAPDPSAFSYPGPH 1082
Db 1009 SKSP-----NRE-----W-----EVLQAPH-----QVITNLPEGV 1034
QY 1083 PLPLGLHDTARPVLPPTPTISNPPLISSAKHPSVLEROIGALSQ-----MSVQLHVPY 1137
Db 1035 RLP-----TTRTPRPPPLIPSKTTVAASEK-PSFI--MGSGISQGTGTYLSSHNOAYP 1086
QY 1138 SEHAKAPGVVWMLPLPMDPKLAPFSGVKQQLSPRGOAGPPELSGVPTAEASVLRG 1197
Db 1087 QEAPKPSVGSISLGLPRQOEKTAAPLTVIKQEFSPRSONSQPEGLLV-RAQHEGVVRG 1145
QY 1198 TALGSPVSGSIYKIGIPSTRVPSDAITYRGSITHTGTPA-----DVLVKGITRI-IG 1248
Db 1146 TA-GAVQESITNGTTPASKISVETISSLRGSIQTGTPALPOAGIPTALVKGVPSRMPIE 1204
QY 1249 EDSPSRLDRGSDSLPKGHVIEGKKGHVLSYSGGMSVTQCSKEDGRSSGPPHETAAPK 1308
Db 1205 ESSPEKV---REEAASKGHVIEGKSGHLSYDNINKNA-----REGTRSPRTAHMSL-K 1255
QY 1309 RTYDMMEGRVGRAS-----SASIEGLMGRATIPPRHSPH-HLKEQHHRIGSITQIGPR 1361
Db 1256 RSYEAVEGSIKQGMRESVPVSAPEGLICRALP--RGSPHSLKERTVLSGSIQGTTPR 1313
QY 1362 SYVEAQEDYLREAKLLKREGTPPPPPPPSPRDLTEAYKTQALGLKPAHEGLVATVKEA 1421
Db 1314 ATAESPEDGL-KYPKQIKRES-----PPIRAFEGAI-----TKGP-YDG-ITTIKEM 1358

QY	1422	GRS	HEIPREEL-----RHTPELAPRPLKEGSI	TQGTPLKYD	TGASTGSKXHDVRS	1476		
DB	1359	GRS	HEIPRODIL	TQESRKTEPVQ	STRPIEGSIQGTPIKEDNN-SQSAIKHNKVS	1417		
QY	1477	IGSP	RTFFPVHLDVWAD-ARALERACYE-----	ESLKR	PGTASSGGS	IARGAPVIV 1530		
DB	1418	ITGPSK	L-----PRGML	LEIVPENIKVREK	YBDVKAGEPVRARHTSVVSSG	SVLST--L 1472		
QY	1531	PELGK	PROSLTYEDHGA-----	PFAGHL	PRGSPVTMR	EPTRLQEGSLSSK-ASODRK 1584		
DB	1473	HEAPK	AGLSCLYDDSSAR	TPVSYQNTISRG	SPMNR-----	TSDVSSSKASHERK 1525		
QY	1585	LTS	TPRE-----IAKSPH	STVPEHPHPISPYEHL	LRGVGVDDLYRSHIP	PLAFDPTSIPRG 1640		
DB	1526	STLT	TORESIPAKSPV	GVGDVIVSH--SPDP	PHRSSAAGEVYRSHL	PTHLPD-AMPFH 1582		
QY	1641	IP	DAAAAYLPH	LAPNPYPHYLLY	TRGVDP	TAALEN-ROTTINDYITSQOMHNT 1599		
DB	1583	RALD	PAAYLLQRLS	TPGSPQYLY-----	AMENT	ROTTINDYITSQOMQVNL 1633		
QY	1700	ATAMA	QADMLRGLSP	RESSIALNYA	AGPRGIIDL	SOVPHLPVLVPT	PGCPATAMDRLA 1759	
DB	1634	-----	RPD	VTIRGLSPRE	QPLGYPYA-TGII	IDLTNMPP-TILVPHAG	STSTPPMDRIT 1685	
QY	1760	YLTA	POQFSSR-HSS	PLSPGPGFTHLTKPT	TTTTSSSERERDRDRDR	DREREKSI	LTSIT 1818	
DB	1686	YI	PGTVTPPRPV	NAASLSPGHPTHL--	AAAA	SAERERERERERERERERER 1742		
QY	1819	TTVEH	AP-----	IWRPGTEOSSGSSGGG	SSRSPASHASHAHQHPIS	PRTODA-LOOR 1874		
DB	1743	ERTAA	PADLYLRPG	SEQ-----	PGR	GSHGVYRSPSP-SVRTQETILQOR 1787		
QY	1875	PSV	LHNTGMKGI	ITAVPSPKPTVLR	STSTSPV-----	RPAATFP	PPATHCPGLGTLDGVYP 1930	
DB	1788	PSV	FGTNGT	SVITPLDPTAQL	RMPLPSGGPSISQGL	PASRYNTAADA-LAALV	DAAS 1846	
QY	1931	TLME	VLLPKA	PR-----	VARPE	PRADTGHAF	LAKPPARSGLEBPASSPK 1977	
DB	1847	APQMD	VS	KTESKHEAAR	LEENLRSAAYSEQQ	LEKQKLEVKRSVQCVC	TSSTALPSG 1906	
QY	1978	GSEPR	PLV-----	PPVSG	HATIA	TPAKN--LAPHAS	POPPAPPASASPHRE 2024	
DB	1907	KAQPH	ASVYVE	AGKOGPPPKRYE	BEELTRGTTITAN	FDIVIT	ITROIASDKDARER 1966	
QY	2025	KTQSK	PSIOE	LELRSLGYHGSSY	SPGVPSPVSPSPSL	THDKGLPKHLEELD	KSHLEG 2084	
DB	1967	GSQ	SDSS-----	SSL	SHRYETASDAIEV	SPASSPA	PQEQPAYQPMVKANQAE 2020	
QY	2085	ELAPK	QGPVKL	GGEAHLPHLR-----	PLPESQ	SPSSPLLOT--APGVK	GHORVTLAQ 2137	
DB	2021	ESTRQ	YEGP-----	LHRYASQ	BSPPSQOQPLP	SPSSQEGMQVPR	HLRLITLAD 2071	
QY	2138	HISEV	ITQDYTRHH-POOL	SAPL-----	PAPLY	SPFGASC	PVLDLRRPPSLYLPDPD-- 2189	
DB	2072	HI	COITQD	FARQVPSQAST	STFTQTS	SPALSSTP-----	VRTKTSRYSPESQSOT 2123	
QY	2190	--	HGAP-RGSPH-----	SEGGK	SRSPFNKTSVL	GGGEDIEP	VSPPEGMEFEG-HSRSA 2240	
DB	2124	VLHPR	PGPRVSPEN	LVKSRGSPK	SPERSHI--PSE	PEPIS	PPQG--PAVHEKQD 2177	
QY	2241	VYPLL	YRDGQTEPS	RMGSKSPGNT	OPPAFFSKL	TESN	AMVKSQKEINKKLNTNRN 2300	
DB	2178	SM	LLSQGVDP	PAEQRSDSR	SPGISY	LPSPFFTKL-ESTSP	WMVKSQKEIFRKLNSGGG 2236	
QY	2301	EPEVNI	QPTETEF	NMPAITGTGLMT	YRSQAVOB	HASTN	GLEAIRKALMGKYDQWEE- 2359	
DB	2237	DSN	WAARQPT	EFNLPAVTT	SGAVSRSH	SFADPAS-NIG	LEDIRKALMG	SFDDKVED 2295
QY	2360	-----	SP	LSANAFNPL	NASAS	UPAAMPIT	AAAGRSHTLITSPGGG-GKAKV	SGRPFSSRK 2413
DB	2296	HGVW	SHPV--	GTMPS	ASTSV-----	VTSS	EARDEGPPSHAGYCKP	KLINKNSRK 2347

Qy	2414	AKSPAPGLA--SGDRPPSVSVSHSEGDCHNRRTPLTNRWVEDRPSAGSTPPFNPPLIMRL	2477
Db	2348	SKSPIPQSGVLGTRPSSSVSHSEGDYHRQTP--GWAWEDRPSSTGSTQFFNPPLTIRM	2405
Qy	2472	QAGVMASPPPPGLPAGSGPL--AGPH--HAWDEEPKPLLCSOVETLSDSSE	2517
Db	2406	-----LSSTPTQTACAPSAITQAPHQONRIWEREPAPLLSAQYETLSDD	2452
RESULT 7			
AAB12453			
ID	AAB12453	standard; protein; 2343 AA.	
XX	AAB12453;		
XX	AC		
XX	AC		
DT	24-OCT-2000	(first entry)	
XX	Human HNRCR	protein SEQ ID NO:20.	
DE	Human; HNRCR;	nuclear receptor coreceptor.	
KW	Human; HNRCR;	nuclear receptor coreceptor.	
XX	XX		
XX	OS	Homo sapiens.	
XX	XX		
FN	CN1250094-A.		
XX	XX		
PD	12-APR-2000.		
XX	XX		
PF	06-OCT-1998;	98CN-00120919.	
XX	XX		
PR	06-OCT-1998;	98CN-00120919.	
XX	XX		
PA	(XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.		
XX	XX		
PI	Yu L, Tu Q, Zhao Y;		
DR	WPI; 2000-400830/35.		
DR	N-PSDB; AAA60629.		
XX	XX		
PT	Preparation of new human kernon acceptor co-repressor coding series and		
PT	the polypeptide.		
XX	XX		
PS	Claim 4; Page 25-30; 58pp; Chinese.		
XX	XX		
CC	The present sequence represents a human homologue of nuclear receptor		
CC	coreceptor (HNRCR)		
XX	XX		
SQ	Sequence 2343 AA;		
Query Match 30.6%; Score 4043; DB 3; Length 2343;			
Best Local Similarity 41.3%; Pred. No. 2.5e-218;			
Matches 1061; Conservative 330; Mismatches 731; Indels 444; Gaps 104;			
Qy	137	KDRSLTGKLE-PVSPSPPHDPELELVPPRLSKEELIQNDRVDRBITWVQOISKLKK	195
Db	36	QDPAFGKGHEAPSSPISGQGDQNASPSKLSKEELIQSDMRVDRIEAKVQOILKXK	95
Qy	196	KQOLEEEAAKPEPEKPVSPPTESKGRSLVQIIVDENRKKAEAAHRILEGPGOVPELP	255
Db	96	KQOLEEEAAKPEPEKPVSPPPVQEKHRSIVQIIVDENRKKAEAAHRILEGPGOVPELP	155
Qy	256	LYNQPSDTROYHENIKINQAMRKKLILYFKRRNHARKQWKQFCORYDQLEALEKKVER	315
Db	156	LYNQPSDTKYHENIKTNQVWRKKLILFFKRRNHARKQEQKICQRYDQLEAWKVKDR	215
Qy	316	IENNPRAKESKVREYVEKQPEIRKQRELQRMOSRVQQRSGLSMSAARSEHEVSEI	375
Db	216	IENNPRAKESKTREYVEKQPEIRKQRBQERFQ-RVQQRGAGLSATTIARSEHISEI	274
Qy	376	IDGLSEQNELEKQMRQLAVIPMLYDAQDQORIKFINMNGIMADPMKYKDVQVNMWMSQ	435
Db	275	IDGLSEQSENEKMRQLSVIPPMFMFDAQRRVXFNNMGLMEDPMKYKDVQVNMWIDH	334
Qy	436	EKETFRFKFQHPKNFGLIASFLERKTVAEKCVLIYYILTKKNYKSLVRSY--RRRGKSQ	494

QY	2476	MASPPPPGLPAGSG-PLAGPH--HAWDEPKPLLCQVETLSDSE	2517
DB	2297	SSTPPTPIACAPSAVNQAAPHQNRIRWEREPAPLISAQVETLSDSD	2342
RESULT 8			
ABU70857			
ID	ABU70857	standard; protein; 216 AA.	
XX	AC	ABU70857;	
XX	DT	10-JUN-2003 (first entry)	
DE	DE	Human adipocyte Selected Interacting domain, SID, #488.	
XX	KW	Human; prey; adipocyte; SID; selected interacting domain; anorectic;	
KW	KW	antidiabetic; protein-protein interaction; diabetes;	
XX	KW	yeast 2-hybrid assay; metabolic disorder; obesity.	
OS	OS	Homo sapiens.	
XX	XX	WO200286122-A2.	
PN	XX	31-OCT-2002.	
PD	XX	14-MAR-2002; 2002WO-EP003768.	
PF	XX	14-MAR-2001; 2001US-0275734P.	
XX	XX	(HYBR-) HYBRIGENICS.	
PA	XX	Legrain P, Daviet L;	
PI	XX	WPI; 2003-103412/09.	
XX	DR	N-PSDB; ACA57401.	
XX	PT	New complex between two interacting proteins in adipocyte cells, useful	
PT	PT	for identifying selected interacting domains that modulate protein	
PT	PT	interactions, or for preventing or treating metabolic disorders such as	
PT	PT	obesity or diabetes.	
XX	PS	Claim 6; Page 271; 382pp; English.	
XX	CC	The invention relates to a complex between two interacting proteins in	
CC	CC	adipocyte cells, given in the specification. The proteins are identified	
CC	CC	by selecting a bait protein from a known adipocyte marker and then	
CC	CC	performing a yeast 2-hybrid selection to isolate prey proteins encoded by	
CC	CC	members of an adipocyte cDNA library. The proteins are designated SID	
CC	CC	(RTM) (selected interacting domains) proteins. Also included are a	
CC	CC	polynucleotide encoding a polypeptide in the adipocyte cells, a	
CC	CC	recombinant host cell expressing at least one of the interacting	
CC	CC	polypeptides of the complex, selecting a modulating compound in adipocyte	
CC	CC	cells, a SID (RTM) polypeptide comprising any of the 738 amino acid	
CC	CC	sequences given in the specification (including its fragment or variant),	
CC	CC	a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences	
CC	CC	given in the specification (including its fragment or variant), a vector	
CC	CC	comprising the SID (RTM) polynucleotide, a recombinant host cell	
CC	CC	comprising the vector, a protein chip comprising the polypeptides and a	
CC	CC	record comprising all or part of the data, listed in the specification.	
CC	CC	The complex, polypeptides, polynucleotides and compounds are useful for	
CC	CC	preventing or treating metabolic disorders such as obesity or diabetes.	
CC	CC	The polynucleotides are useful as probes or primers. The complex is	
CC	CC	particularly useful for identifying selected interacting domains (SID	
CC	CC	(RTM)) for screening drugs that modulate the protein interaction, thus	
CC	CC	exhibiting the therapeutic effect. The present sequence represents a SID	
CC	CC	(prey) protein of the invention	
XX	XX	Sequence 216 AA;	
Query Match			
Best Local Similarity 8.4%; Score 1113; DB 6; Length 216;			
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Query Match			
Best Local Similarity 8.0%; Score 1060; DB 4; Length 3502;			
Matches 699; Conservative 346; Mismatches 1056; Indels 1234; Gaps 137;			
QY	20	PPHSLSYVQIARTHTDVGLEYQH--HSRDYASHLSPGSIQPORRRPS-----	67

QY	203	EAAPPEPEKPVSPPIESKHSLSLQIYYDENRKAHAHRILEGIGPOVELPLYNQPSD	262
DB	1	EAAPPEPEKPVSPPIESKHSLSLQIYYDENRKAHAHRILEGIGPOVELPLYNQPSD	60
QY	263	TRQYHENIKINQAMRKLLILYFKRRNHARKQWKQFCQRYDQJMEALEKKVERIENPPRR	322
DB	61	TRQYHENIKINQAMRKLLILYFKRRNHARKQWKQFCQRYDQJMEALEKKVERIENPPRR	120
QY	323	RAKESKVREYIEKQPEIRKQRELQERMQSRVQSGSLGMSAARSEHVSIIIDGLSEQ	382
DB	121	RAKESKVREYIEKQPEIRKQRELQERMQSRVQSGSLGMSAARSEHVSIIIDGLSEQ	180
QY	383	ENLEKQMRQLAVIPPMLYDADQORIKFINNGLMAD	418
DB	181	ENLEKQMRQLAVIPPMLYDADQORIKFINNGLMAD	216
RESULT 9			
ABBS8382			
ID	ABBS8382	standard; protein; 3502 AA.	
XX	AC	ABBS8382;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 1938.	
XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
KW	KW	pharmaceutical.	
XX	OS	Drosophila melanogaster.	
XX	PN	WO200171042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-US009231.	
XX	PR	23-MAR-2000; 2000US-0191637P.	
XX	PR	11-JUL-2000; 2000US-00614150.	
XX	XX	(PEKE) PE CORP NY.	
PI	PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	DR	WPI; 2001-656860/75.	
XX	DR	N-PSDB; ABL02485.	
XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	PT	genes from Drosophila and for elucidating cell signaling and cell-cell	
PT	PT	interactions.	
XX	PS	Disclosure; SEQ ID NO 1938; 21pp + Sequence Listing; English.	
XX	CC	The invention relates to an isolated nucleic acid detection reagent	
CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
CC	CC	useful in developmental biology and in elucidating cell signalling and	
CC	CC	cell-cell interactions in higher eukaryotes for the development of	
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
CC	CC	sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-	
CC	CC	ABB72072). The sequence data for this patent did not form part of the	
CC	CC	printed specification, but was obtained in electronic format directly	
CC	CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX	XX	Sequence 3502 AA;	
Query Match			
Best Local Similarity 21.0%; Pred. No. 4e-50;			
Matches 699; Conservative 346; Mismatches 1056; Indels 1234; Gaps 137;			

QY 138 DRSITGKLEPVSPSPHTDPELEVPRLSKELIQNMMDRVREITMVQOISKLKKQ 197
DB 1 DRSITGKLEPVSPSPHTDPELEVPRLSKELIQNMMDRVREITMVQOISKLKKQ 60
QY 198 QOLEEAAKPEPEKPVSPPIESKHSLSVQIYDENRKAEEAHRILEGLGQVQLPLY 257
DB 61 QOLEEAAKPEPEKPVSPPIESKHSLSVQIYDENRKAEEAHRILEGLGQVQLPLY 120
QY 258 NQPSDTRQYHENIKINQAMRKKILYFKRRNHARKQWKQFCQRYDQLEALEKKVERIE 317
DB 121 NQPSDTRQYHENIKINQAMRKKILYFKRRNHARKQWKQFCQRYDQLEALEKKVERIE 180
QY 318 NNPR 321
DB 181 NNPR 184
RESULT 11
ABU70979
ID ABU70979 standard; protein; 224 AA.
XX
AC ABU70979;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human adipocyte Selected Interacting domain, SID, #610.
XX
KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;
KW anti-diabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS Homo sapiens.
XX
PN WO200286122-A2.
XX
PD 31-OCT-2002.
XX
PF 14-MAR-2002; 2002WO-EP003768.
XX
PR 14-MAR-2001; 2001US-0275734P.
XX
PA (HYBR-) HYBRIGENICS.
XX
PI Legrain P, Daviet L;
XX
DR WPI; 2003-103412/09.
DR N-PSDB; ACAS7523.
XX
PT New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes.
XX
PS Claim 6; Page 308-309; 382pp; English.
XX
CC The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and a
CC record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are useful for
CC preventing or treating metabolic disorders such as obesity or diabetes.

CC The polynucleotides are useful as probes or primers. The complex is
CC particularly useful for identifying selected interacting domains (SID
CC (RTM)) for screening drugs that modulate the protein interaction, thus
CC exhibiting the therapeutic effect. The present sequence represents a SID
CC (prey) protein of the invention
XX
SQ Sequence 224 AA;
Query Match 5.9%; Score 774.5; DB 6; Length 224;
Best Local Similarity 73.6%; Pred. No. 1.3e-35;
Matches 159; Conservative 14; Mismatches 38; Indels 5; Gaps 3;
QY 116 PLLRSPPLATGQAGSEDLTKDRSLTGKLE-PVSPSPHTDPELEVPRLSKELIQ 174
DB 13 PLVHPLP---EGLRA-SADAKKDPAGFGKHEAPSSISQPCDDQNASPSKLSKEELIQ 68
QY 175 NMDRVREITMVQOISKLKKQOOLEEAAKPEPEKPVSPPIESKHSLSVQIYDEN 234
DB 69 SMDRVREITAKVEQQILKUKKQOOLEEAAKPEPEKPVSPPIESKHSLSVQIYDEN 128
QY 235 RKKAEEAHRILEGLGQVQLPLYNOPSDTROYHENIKINQAMRKKILYFKRRNHARKQ 294
DB 129 RKKAEEAHRILEGLGQVQLPLYNOPSDTROYHENIKINQAMRKKILYFKRRNHARKQ 188
QY 295 KQKFCQRYDQLEALEKKVERIENNRRAKESKVR 330
DB 189 EQKICQRYDQLEALEKKVERIENNRRAKESKTR 224
RESULT 12
AAR99738
ID AAR99738 standard; protein; 619 AA.
XX
AC AAR99738;
XX
DT 27-SEP-1996 (first entry)
XX
DE Retinoid X receptor interacting protein RIP13.
XX
KW Retinoid X receptor interacting protein; RXR; RIP; RIP13.
XX
OS Mus sp.
XX
PN WO9621677-A1.
XX
PD 18-JUL-1996.
XX
PF 08-DEC-1995; 95WO-US016311.
XX
PR 13-JAN-1995; 95US-00372652.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Moore D, Seol W, Choi H;
XX
DR WPI; 1996-342241/34.
DR N-PSDB; AAT31931.
XX
PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to modulate
PT or mediate RXR function, anti-RIP antibodies can be used to determine RIP
PT subcellular distribution patterns.
XX
PS Claim 2; Page 51-52; 90pp; English.
XX
CC Mouse retinoid X receptor (RXR) interacting protein RIP13 (AAR99738) is a
CC candidate transcriptional co-activator. It was identified using an in
CC vivo interaction trap system for the isolation of proteins that
CC physically interact with RXRs, esp. with the ligand binding domain of
CC human RXR alpha. Recombinant RIP13 can be obtd. using a cDNA clone
CC (AAT31931) obtd. from a mouse liver library. RIFs (see also AAR99735-37
CC and AAR99739) can be used to modulate or mediate RXR function, and may be
CC used therapeutically or to raise antibodies
XX

SQ	Sequence 619 AA;	
XX	Query Match	
XX	Best Local Similarity 5.4%; Score 718; DB 2; Length 619;	
XX	Matches 213; Conservative 76; Mismatches 212; Indels 90; Gaps 26;	
QY	1975 PSKGSRRPLV-----PPVGHGATIAITPAKN--LAPHHASDPAPPASADP 2021	XX (EXEL-) EXELIXIS INC.
Db	70 PSQAQPHASVYVSEAGKDGPPKRYEELTRGKTTITAAAFIDVITRIQIASDKA 129	XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
QY	2022 HREKTSQKPSIQEELRLSLGVHGSYSPGVEPVSPVSPSLTHDKPLKHELDKSH 2081	XX WPI; 2003-156859/15.
Db	130 RERGSSDSS-----SSLSHRYETASDAIEVISPASSAPPOEKPOAYOPDMVKANQ 183	XX N-PSDB; ACD13385.
QY	2082 LEGELRPQGPVKLGGAHLPLR-----PLPESQSSSPLLOF--APGVKGHORVVT 2134	XX Identifying modulators of the p53 pathway for use in treating apoptotic
Db	184 AENESTROVEGP-----LHYSQOESPSPOQPLPSSQSEGQGVPRTHRLIT 234	XX or cell proliferation disorders, comprises screening for agents that
QY	2135 LAQHISEVITQDYTRHHPQOLGAPLPAPLYSPFGA--SCPVLDRRPPSDLYLPPPD--- 2189	XX modulate activity of a human ortholog of genes that modify the p53
Db	235 LADHICQIITQDFARN--QVPSQPSSTQTSPSALSSTPV---RTKTSRYSPEQSQT 289	XX pathway in Drosophila.
QY	2190 --HGAPA-RGSPH-----SEGGKRSPEPNKTSVLGGEDGIEPVSPGWTBPG-HRSA 2240	XX Example 2; Page 505-515; 678pp; English.
Db	290 VLHPRGPRVSPENLVDKSRGSRPGKSPERSHI---PSBPYEPISPPQG---PAVHEKQD 343	XX The invention relates to identifying (M1) a candidate p53 pathway
QY	2241 VYLLYRDGEOTEPSRMGSKSGNTSQPPAFTSKLTESNAMYKSKOEINKLANTHNN 2300	XX modulating agent, by contacting an assay system comprising a purified HM
Db	344 SMLLSQRGVDPAEQSDRSRPSGISYLPFFTKL--ESTSPWKSGKQEIFRKINSGGG 402	XX polypeptide (human orthologue of genes that modify the p53 pathway in
QY	2301 EPEYNTSQPQTEIFNPAITGTGLMYRSQAOEHASTNMGLEAITRKALMGKYDQWEE- 2359	XX Drosophila) or nucleic acid with a test agent under conditions, where but
Db	403 DSDMAAQCTEIFNPAVTTGAVSSRSHSPADPAS--NLGLEDIIRKALMGSDDKVED 461	XX for the presence of the test agent, the system provides a reference
QY	2360 -----SPPLSANAFNLINASALPAMPITAADGRSDHILTSPGGG-GKAKVSGRPSRK 2413	XX activity, and detecting a test agent-biased activity of a cell (comprising
Db	462 HGVVMGHPV---GIMFGSASTSV-----VTSSEARDEGEPSPHAGVCVKPKLINKNSRK 513	XX Also included are modulating (M2) a p53 pathway of a cell (comprising
QY	2414 AKSPAPGLA--SGDRPPSVSVHSEGDCHNRRTPLTNRVWEDRPSAGSTPPFVNPILMRL 2471	XX contacting a cell defective in p53 function with a candidate modulator
Db	514 SKSPIQSGVLGTERSSSVSVHSEGDYHKQTP--GWAMEDRPSSTGSTQFPNPLTIRM 571	XX that specifically binds to a HM polypeptide comprising an HM amino acid
QY	2472 QAGVMASSPPPPGLPAGSGPL--AGPH---HAWDEPKPLCSQYETLSDSE 2517	XX sequence, where p53 function is restored), modulating (M3) a p53 pathway
Db	572 ----LSSTPTQTACAPSAITQAAPHQONRIWEREPAPLLSAQYETLSDSD 618	XX in a mammalian cell (comprising contacting the cell with an agent that
XX		XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
XX		XX a disease in a patient (comprising: (a) obtaining a biological sample
XX		XX from the patient; (b) contacting the sample with a probe for HM
XX		XX expression; (c) comparing the results with a control; and (d) determining
XX		XX whether the comparison indicates a likelihood disease). (M1) is useful
XX		XX for identifying modulators of the p53 pathway. A probe for HM expression
XX		XX is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
XX		XX in a patient, where the cancer has greater than 25 % expression level.
XX		XX Modulators identified by (M1) are useful in a variety of diagnostic and
XX		XX therapeutic applications, where disease or disorder prognosis is related
XX		XX to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
XX		XX proliferation disorders (e.g. cancer). Another two new methods (M2 and
XX		XX M3) are useful for modulating the p53 pathway of a cell, thus restoring
XX		XX the p53 function of the cell, so that the cell undergoes normal
XX		XX proliferation or progression through the cell cycle. (M2) and (M3) are
XX		XX also useful for treating defects in the p53 pathway such as angiogenic,
XX		XX apoptotic or cell proliferation disorders. The present sequence
XX		XX represents a human p53 pathway modifying protein
SQ	Sequence 3371 AA;	
QY	71 EFQPGNRSQELHLRPE-----SHSLPELKGK-----EMETESKR---PRLELPL 114	XX Query Match
Db	849 ERKSGQKSHSVNTEKIGIDIDHTQSYRKQKQRRKQKQMEETAKSKFSPKDV-- 906	XX Best Local Similarity 4.3%; Score 569.5; DB 6; Length 3371;
QY	115 DPLLRPSPLLATCQAGSEDLTKDRSLTKGLBFVSPSPPPHTDPELELVPRLSKEELIQ 174	XX Mismatches 568; Conservative 349; Mismatches 973; Indels 1011; Gaps 137;
Db	907 DEVERSLVHEVGKP--PODVTDD-----SPPSKK-----K 935	XX
QY	175 NMDRVDRDI--TWVQOISLKKKKQOOLEEAAKPP-----EPEKVPSPPIES----- 221	XX
Db	936 RMDHVDICTKRENRYSRQISEDSERTGSGSPSVRHGSHFDEDEPICSPLLSVKGSP 995	XX
QY	222 -----KHSLSVLQIYDENRKK--AEAHRILEGLGPOVELPLYNQPSDTRQ 265	XX
Db	996 KYDEKVLPSYNTVREESLKFNPYSSRREQADWAKIKLSVLNSEDENRW-----DSQM 1051	XX
QY	266 YHENTKINQAMRKLLIYFKRRNHAKQWKQFCORYDQLMFALEKKVRIENNPR-- 322	XX
Db	1052 KQDAGRFDVSPFNSII-----KDSLRLKRSVRDL--EPGEVPSDSDGDEGSHSPRASAL 1105	XX
QY	323 -----RAKESKVREYEEKQPEIRKRELQERMSRVQGRSGL-----SNSAAR 367	XX

QY 2196 GS-----PH-----SEGGKRSPEPNKTSVLGGEDG 2221
Db 2882 SEVLVMQSEYRLHPYTPVDRVIMVHPHTAVSEQRAADGVVKKVPASKAP-----QQPG 2937
QY 2222 IEVSPPEGHTPECHRSAY-----YPLLVDGEQTPESRMGSKSPGNTSQ-----PP 2269
Db 2938 KEAAKTPDAKAAPTPTAPVPVPLPAPAPAPHGE-----ARILVTSPNQLQGLPLTPP 2993
QY 2270 APFSKLTESAMVKKQKQINKLKNHNRNE--PEYNI SQPCTEIFNMPAITGTGLMTY 2327
Db 2994 -----VVVTHGVQI-----VHSGGELFQYRYGD-----IRTY 3021
QY 2328 RSOAQOEH-----ASTNMGLEAIRKALMGKYDOWEE--SPPLSANAFNPLNASASLPAAMP 2382
Db 3022 HPPAQLTHQTQFPAASSVGLPSRTKTAQAQPPPEGEPLQPPQVQSQPAQAPPCPPSQ- 3080
QY 2383 ITAADGRSDHLLTSPGGGKAK---VSGRPSSRKAKSPAGLASG---DRPP-----SVSSV 2433
Db 3081 -----LQPGQGPFSKMPQVSOEAKGTGTGVEQFRLPAGPANRPPEFHTQVQRA 3129
QY 2434 HSB-GDCNRRTPLTNRVWEDRPSAGSTPPFYVPLIMRLQAGVMASPPPPGLPAGSGPLA 2492
Db 3130 QAGTGTSPSPVSVSMKDLPLVSLFTQAPKQPLFVPTTSG-----PSTPPG---LV 3179
QY 2493 GPHAWDEEPK-----ELLCSSQ 2509
Db 3180 LPHTFQAPKQDSSPHLTSQ 3200

RESULT 14

ABR47592
ID ABR47592 standard; protein; 3664 AA.

AC ABR47592;

XX 12-JUN-2003 (first entry)

DE Breast cancer associated protein sequence SEQ ID NO:423.

DE Human; breast cancer; cytostatic; gene therapy.

XX Homo sapiens.

OS WO2003004989-A2.

PN 16-JAN-2003.

XX 21-JUN-2002; 2002WO-US019669.

XX 21-JUN-2001; 2001US-0299887P.

PR 27-JUN-2001; 2001US-0301572P.

PR 18-JUL-2001; 2001US-0306501P.

PR 25-SEP-2001; 2001US-0325002P.

PR 05-MAR-2002; 2002US-0362585P.

PR 14-MAY-2002; 2002US-0380391P.

XX (WILL-) MILLENIUM PHARM INC.

XX Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kanatkar S;

PI Mertens M, Monahan JB, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;

PI Bast RC, Hortobagyi GN, Pusztai L, Mexic F, Sahin A, Mills GB;

XX WPI; 2003-210381/20.

DR N-PSDB; ACC50291.

XX Breast cancer diagnosis or treatment by comparing the level of expression

PT of a marker in a patient sample with that in the control non-breast

PT cancer sample.

CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences from
CC the present invention have cytostatic activities and can be used in gene
CC therapy. The method is useful for diagnosing and treating breast cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3664 AA;

Query Match 4.3%; Score 569.5; DB 6; Length 3664;

Best Local Similarity 19.6%; Pred. No. 1.8e-22; Indels 1011; Gaps 137;
Matches 568; Conservative 349; Mismatches 973;

QY 71 EFQGNERSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR---PRLELLP 114
Db 1142 ERKSQEKSHSVNTEKIGIDIDHTQSYRKQMEQSRKQOMEMETAKSEKFGSPKQDV-- 1199
QY 115 DPLLPPSLATGQAGSSEDLTKDRSLTGKLEPVPSPPPHTDPELELVPPRLSKEELIQ 174
Db 1200 DEYERRSLVHEVGKP--PDVTTD-----SPPSKK-----K 1228
QY 175 NMDRVDRET--TWVEQOISKLKKQQLSEAAKPP-----EPEKVPSPPIES----- 221
Db 1229 RMDHVDFDICTKRNRYSSROI SEDSERTGSGSVRHGSHFHEDEDFGSPRLLSVKGSP 1288
QY 222 -----KHSLSVQIYDENRKK--AEAHRILEGLQPQVELPLYNQSDTRQ 265
Db 1289 KVDEKVLPSYNTVREESLKNFYDSRRREQWADWAKIKLSVLNSEDELNRW---DSQM 1344
QY 266 YHENIKINQAMRKILLYFKRRNHARKQWKQFCQRYDQLMALSKKVERIENRPR--- 322
Db 1345 KQDAGRFVDFVFNFI-----KRDLSLRKRSVRL--EPGEVPSDSDEDEGHHKSHSPRASAL 1398
QY 323 -----RAKESKRVREYKQFPEIRKQRELOERMOSRVGQSGI-----SMSAAR 367
Db 1399 YESSRLSFLLRDREKLRDERLSLSLERNKFPYFALDKTITPTDKALLERAKSLSSSR 1458
QY 368 SEHVESEIIDGLSE---QENLEKQWQLAVIPPLMYDADQQRKIFINMGLMADPMKYV 423
Db 1459 EEN--WSFLWDSDRFANFRNNKDEKVD SAPPIPSWYMKKKIR--TDSEGRWMDKKEDH 1515
QY 424 KDRQVNMWSEOEKETFREKFMQHPNFGLTASFLEKRTVAECVLYYLTKKNENYKSLV 483
Db 1516 KEE-----QERQLFASRL--HSSI FEQDSKRLQ-----HLERKEEDSDFTS 1557
QY 484 RRSYRRRGKSQQOQQOQQOQQOQQOQQOQ-----MPRSSQEKDEKEKEKEKEE 533
Db 1558 GRIY---GK---QTSSEGANSTTDSIQEPVVLPHSRFMELTMQQKEKEKDKQPKVEKQE 1611
QY 534 EKPEYENDKEDLLKEKTDGNDENDEKAEVASKGRKTANSQGRKGRITRSMANEANSE 593
Db 1612 ---DTE-----HPKTPESAPENKD-----SELKTP 1634
QY 594 EAITQQQAELASLMELNESSRWTEEMETAKGLLEHGNWNSAIARMVGSKTIVSOCKNFP 653
Db 1635 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1663
QY 654 FNYKQRQLDEILQHLKMKERNARRKKKAPAAASEAAFPVWDEEEMEAASVSGN 713
Db 1664 -----EAPLVTEKTVEPATVSEAKPASEPAPA 1692
QY 714 EEMVVEEAALHASGNEVPRGCSGPATVNNSSDTEISPSHTEAAKDTGQNGPKPPATL 773
Db 1693 PVEQLEQV-----DLPPGA-----DPDKAAAMP--AGVEEGSGDQPP-YL 1731
QY 774 GADGPPGPPPPPPRRTSRAP--IEPTASEATGATPPPPAPPSPSPAPPVVPVKEKEEET 831
Db 1732 DAK-----PPTPGSFSQAESNVDPEDPS-----TQPLSKPAQKSEANEPKPKPDAT 1780

CC V(D)J Res) activity or that modulates the interaction between KRC and a
 CC TRAF molecule, so that inflammation or apoptosis in the cell is
 CC modulated. Also provided are methods for modulating immune cell
 CC proliferation or activation and for inhibiting metastatic growth of a
 CC tumour cell. The method is useful in modulating an immune response, such
 CC as inflammation or apoptosis, using agents that modulate KRC activity.
 CC The method may also be used in identifying agents that modulate KRC
 CC activity which can be used for treating or preventing disorders
 CC associated with an aberrant cell proliferation or survival, such as
 CC autoimmune disorders or graft-versus-host disease. The present sequence
 CC represents a human KRC protein, a DNA binding protein. KRC is a member of
 CC zinc finger proteins that bind to the kappaB motif
 XX
 SQ Sequence 2406 AA;

Query Match 4.2%; Score 552.5; DB 6; Length 2406;
 Best Local Similarity 19.8%; Pred. No. 9.2e-22;
 Matches 553; Conservative 323; Mismatches 966; Indels 953; Gaps 136;

QY 3 GSTQLVAQTWRATEPRYPHSLSYVQIARTHTDVGLEYQHHSRDYASHLSPGS---II 59
 DB 69 GSOEKTGQ--QOKPKRPPIEASVHI-----SHVPOHPLTPAFMSPCKPEHLEGSTWQLV 122
 QY 60 QQRRRP--SLISEFOQG--NERSQELHLRPEHSYLP-----ELGKS 98
 DB 123 SPWRLGPSGLLA--PGLHPOSQ---LLPFSHASIIPEDLPGVPKVFVPRPSQVSLKPT 176
 QY 99 EMEFTESKPR-----LELLPDPLRPSFLA-----TGO-----PAGESDLTKD--- 138
 DB 177 EBAHKKERPKQPKYICQYCRPCAKPSVLQKHRSHTGERPYPCGPGCFKTKSNLY 236
 QY 139 -----RSLTGKLEPVPSPPHDPPELELVPPRLSKEELIQNMDRVDREITWVEQOI 190
 DB 237 KHRKSHAIRIKAGLASGMGMYPH--GLEMERIPGEFEPEP-----TEGESTDSEET 288
 QY 191 SKLKKKQOOLEEBAAPPEKPVSPPIES--KHRSVLQIYDENRKAERAAHRIEGL 248
 DB 289 SATSGHPAELS-----PRKQPLSSGLYSGSHSSHERCSLSQSTAQSLIEDPPPFV 342
 QY 249 GPQVELPLYNQPSDTRQYHENIKINAMRKLI-----LYEKRNRHAK 292
 DB 343 EFSSEHPLSHKPEDHTTIQKIALRLSERKKWIDEOAFSPGSKGSTESGYFSRESAEQ 402
 QY 293 QMKQKFCQRYDQLEALEKVERIENNRPRRAKESKVREYIEKQFPEIRKQRELOERMOS 352
 DB 403 Q-----VSPPTNAK-----SYAEIIFGKC-----G 423
 QY 353 RVGORGSLMSGAAR-----SEHEVS-----EIIDGLSEQENLEKQMRQLAVIPP 397
 DB 424 RIGQRTAMLTATSTQPLLPSTEDKPSLVPVPTQVIEHTIKLTINE-----AVVDT 478
 QY 398 MLYDADQORIKFTNNGLMADPMVKYKDRQVMNMSEQEKETFREKFMQKNFGLIASF 457
 DB 479 SEIDSVKPRRSLRRSSMESP--KSLYREPLSGHSEKTKPEQSLLSLQHPSTAPPVPL 537
 QY 458 LERKTV--AECVL-----YYLTCKENYKSLVRRSY-----RRRGKSOQOQOQOQ 502
 DB 538 LASHMPSAACTISTPHFPFRGSYSPDDHITDSEALSRSSHVFTSHPRMLKQPAIELPL 597
 QY 503 OQOQOQOQPMPSRQOEKE 562
 DB 598 GGEYSSEEPGSS---KDTASKPSD-----EVE--PKSEELATKKT----- 632
 QY 563 AVASGRKKTANSGRKG-----RITRSMANSEEAITPOOSAEIASMELNESSRWTE 617
 DB 633 ----KGLKT-----KGVIEYECNICGARYKRDNYEAHKYCYSELQAKPISAGTHTS 682
 QY 618 EEMETAKGLLEHGRNWSAIRMVGSKTYSQCKNFYNYKKRQNLDEILQOHLKMEKER 677
 DB 683 PE---AEKQIEH--EPWSQ-----MMHYKLGJTTLEL 709
 QY 678 NARRKKKAPAAASEAAEPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGEC 737

DB 710 TPLRKRKRKESIGDEE--PPAFESTKQFGS-----P.740
 QY 738 GPATVNNSSDTESIPSPHTEAAKDTGQPKPPATLGADGPPGPPPTP-----PRRT 789
 DB 741 GFSDAARNLPLESTKSP--AEPSKSV-----PSLEGTGFOPRTPKPGSGESGKERT 792
 QY 790 SR---APIEPTPASBATGA-----PTPPAP-----PS 814
 DB 793 TSKEISVIQHTSSFEKSDSLEQPSGLEGDKPLAQFPSPPPAPPHGSAHSLQPKLVQRQN 852
 QY 815 PSAPPVVPKEKEKETAAAPVVEGEOKPPAAABELAVDTCKABEPVKECTEE---ABE 872
 DB 853 IQVPEILVTEEDREDTEPEPPKPEK-----TEEFQWOPORSQTLAQLPASK 900
 QY 873 GFAKGD---AEAAEATAEGALKAEKGGSGRATTAKSSGAPOQSDSATCSADEVDEA 929
 DB 901 APPKKRRLRLAEMAQSGSESSPVP---LSRSPQESNVSLSGSSRSASFERDDHGKA 957
 QY 930 EGGDKNRLSPRP---SLLTPTGDP---RANASPOKP-----LDLKLKQRAAAIIP 975
 DB 958 EAPDPSSDMRPKPLGTHMLTVPSHHPHAREMERSASEQSPNVSHSAHMTETRSKFDYGS 1017
 QY 976 IQVTKVHPPREDAAATPKAPPAP---PPQVLO-----PESD---AP--QO 1014
 DB 1018 LSLT-----GFSAPAVPAPAGEAPPERKCFVRSFSLSRPPESELEVPKGRQ 1067
 QY 1015 PGSSPRGSRSPAPPADKEAFAAEAQKLPDGPCTWGLPPP---VPRPRVKA--- 1065
 DB 1068 ESEEPQSSSKPSAKSLSQISSAATSHGGPGGPGQDRPALGPTVPTVYALQVFFHP 1127
 QY 1066 -----SPHAPDP--SAFSA-----PPGH-----PLPLGLHD 1090
 DB 1128 VAQTLHEKPYLPVPVSLFSFOHLVQHPGQSPPEFSTQAMSSLLSSPYMPPPLPSLFQ 1187
 QY 1091 TARPVLPRPTISNP-----PLISAKHPSVLEROI---GALSQK----- 1129
 DB 1188 A--PPLPLQPTVLPQGLHLPQMPHPANIIPROPFSLPMPYPTSSALSSFFFLPQSO 1245
 QY 1130 -----SVQLHVPYSEHAKAPVGTWGLPLPMDPKKLAPPSPGVKQE--QLSPRGQAGP 1181
 DB 1246 FALQPLPGDVESHLPOIKTSLAPLATGSAG-----LSPQOEYSSDIRLP---VAP 1293
 QY 1182 ESLGVPTAQEASVLRGTALGVPGGSITKGIP---STRVPSDAITYRG-----SITHG 1232
 DB 1294 ASSAAPTAPP---LALPACPDWVSLVVRVQTNMPSYGSAMYTTLSQLVTQSOQ 1348
 QY 1233 TPADVLY-----KGTITRIIGED-----SPSELORGREDSLPKGH---VIYEGK 1274
 DB 1349 SSATVALPKFEPSPSKG---TTVCADVHEVGPSPGLSEEQSRAPPTPYLRVPVTLPERK 1406
 QY 1275 GHVLSEVGMSVTQCSKEDGRSSSGPPHETAPKRTYDM-----MEGRVGRASIASIEG 1329
 DB 1407 GTSLSSESITSL---EGSSSTAGSKXVLSFAGSLELTMTETQOKRVKEEASKADEK 1461
 QY 1330 LMGRAIPP-----ERHSPHLKQHHIRSGITQGIIPRSVVEAQEDYLRREAKLL 1378
 DB 1462 L--ELVKPCSVLVTSTEDGKRPEKSHL-----GNQOG-----RRELEML 1499
 QY 1379 K-----RECTPPPPPSRDLTEAYKTOALGPLKL---KPAHGLVATVKEAGRSIHE 1427
 DB 1500 SSLSDSPDSTKEIPPLPHPALSHGQAPGSEALKEYPQSGKPHRRGLTLP-----S 1550
 QY 1428 IPRESLRHTPELP--LAP---RPLKEGSIQTQTPKYDTGASTTGSKKHVDVSLIGSPGT 1483
 DB 1551 VKKESKQPDPLSLAPSSLPSE---TSSRPAKSQEG---TDSKK-----VLQ 1594
 QY 1484 FPPVH-----PLDVM--ADAPALERA---CYEESLKSRPGTASSGGSIARGA 1526
 DB 1595 FPSLHTTNNVSMCYLNIKPNHIOHADRSSVYAGWCISLNPPLPGVSTKAALSLLR-- 1652
 QY 1527 PVIVELGKPROSLTYEDHGAPFAGHLPGRSPVTWREPTPELOE---GSLSSSKASQDR 1583
 DB 1653 -----SKQVSKETYTWATAP---HPEAGRLVSSSRKPRNTEVHLPLSLVSPGQKDL 1702

```
QY 1584 KLTSTPREIAKSPHSTVP---EHPHPISPYEHLRGVGVLDYRSHIPLAF-----1632
Db 1703 ARVEKEERERGEEDAPASQORGEPAKIFE-----GG---YKSNEEYVYVRGRGRGK 1753
QY 1633 -----DPTSTIPRGIPLDAAAYVLPRL-----APNTYPLHYPPYLIRGYPDT 1676
Db 1754 YVCEBGICRCKPSMLKHIIRTHDVRVYVYKCHFAFTKGNLTJHKMKSKAHKKCQET 1813
QY 1677 AALENRQTIINDYITSQMHNTATAMAQRADMLRGLSPRESSLALNAAAGPRGIIDLQ 1736
Db 1814 GVLE-----ELEAEBGTS-----DDLFDQSEGESEAV-----1842
QY 1737 VPHLPVLVPTPTGPTATAMDRLAYLTAPOPFSRRHSSPLSPGGPHTLTKPTTTSSSR 1796
Db 1843 -----EHHQFSDL-----EDSDS 1855
QY 1797 ERDRERDRDRERKSILTSITT-----VEHAPIWRP-----GTEQSSG 1836
Db 1856 DSDLDDEDEDEESQDELRSRSEAPPGPPhALRADSPILGPPDPAPASGTEATRG 1915
QY 1837 SSGS-----SGGGGSSSRPA-----SHSHAHQHSPISGR-----1866
Db 1916 SVSEAEERLTASSCSMSSQMFGLPWLGPAPLGVSVEKDTGSALSYPVSPRRPWSPSKEA 1975
QY 1867 -TODALQORPSVLHNTGMKIITAVEPSKPTVLIRSTSTSPV-----1907
Db 1976 GSRPPLARKHSLTKNDS-----SPQRCSAPAREPOASAPSPGLHVPDGRGMGLPCG 2027
QY 1908 RPAATFPPTHCPGL-----GTLGVYPTLMEPVLLPKEAP-----RVA 1946
Db 2028 SPRLQSLPTLCPGLRELAAPRAHVLSKLEGTDT---PG-----LPRYSPTRRWSPGQAE 2078
QY 1947 RPERPRADTGHAPLAKPAPARSGLPASPSSKSGSE-----PRPLVPVVS-GHATARTP 1998
Db 2079 SPPR-----SAPPKQWALAGPSPSAGHGGLGAPLVLPFPAPLPHKLSRSP 2128
QY 1999 AKNLAPHASPDPPAPASADPHREKTOSKPFISOELRLSLGYHGSSYPSEGVPSVP 2058
Db 2129 ETCASPMQKA-ESRSPSCSPGAH--PLSSRPFS---ALHDPHGHLARTENI-----2176
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGELRPQKGP-VKLGGEAAHLPHLRPLPESOPSS 2117
Db 2177 -----FSH---LPLHSQHLTRA-----PCPLPIPG-----IQMVQARPGA 2210
QY 2118 PLLQATAP-----GVGQHRVVTLAOHISEVITQDVTTRHHPQOLSAPLPAP---LYSEFG 2168
Db 2211 PTLPGPTAAWVGFGGGSDLTGARE-----AQERGRWSPTESSASVSPVAKVSKFTL 2265
QY 2169 ASCPVLDLRRPPSDLYLPPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGIEFVSP 2228
Db 2266 SS--ELEGDRDYPKERERTGGGPRPDWTPHGTAPEPTFTHS-----PCTPP 2312
QY 2229 EGMTEFCHRSASVPLYLDYDGEOTFSRMGSKSPGN---TSQPPAFPSK-----LTSNSA 2281
Db 2313 DTLPREPQGR-----RAAQWSRPLSPRAPANPEFSATPPLDRSSSVGCLAEA-SA 2363
QY 2282 MVKSKQEIKNKLNTH-----NRNEPEVNIQOP 2309
Db 2364 RPPARTRNLSGESRTRQDSPKPSGSGEPRAHPHQ 2398

RESULT 18
ABB62819
ID ABB62819 standard; protein; 1963 AA.
XX
AC ABB62819;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 15249.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
```

```
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL06922.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 15249; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABLO1840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1963 AA;
```

```
Query Match 4.2%; Score 550; DB 4; Length 1963;
Best Local Similarity 19.9%; Pred. No. 9.7e-22;
Matches 486; Conservative 233; Mismatches 824; Indels 898; Gaps 110;

QY 343 QRELQERMOSRYGQSG---LSMSAARSEHEVSEIIDGLSEQENL-----EKQMRQLAV 394
Db 20 ERELESRRS-PGVADGDLMLFLRAARSMAAFQGMCDGLEDGCLAAASRDDTTINALDV 78
QY 395 IPEMLYDADQORIKFINMGLMADPMKYVKDQVMMWSEQEKETREKPMQHPKPGLI 454
Db 79 LHDSGYDPGKA-----LQALVKCPVSKGIDKK---WTEDETKFKIKGLRQFGKNFFRI 128
QY 455 -ASFLEKRTVAECVLYYLTKNENYKSLVRSYRRGKSOQOQOQOQOQOQOQOQOQMP 513
Db 129 HKDLLPHKDTPELVFYVYVWKTPGANN--NRPHRRRQSGALRRNRVTRANNNSVTP-- 184
QY 514 RSSQEEKDEKEKEAEKEEKEPEVENDEKDLLEKXTDDTSGEDNDEKEAVASGKRKTAN 573
Db 185 -----PKKEDTPEQT-----ATTATAATAA 206
QY 574 SQGRKGRITRSMANPANESEAITPOOSAEALSMELNESRWTEEE-METAKKGLLEHGR 632
Db 207 SE-----TASRSSPAVSKE-----ENSLTEDDASECDSDSLTHKR 243
QY 633 NWSAIARMVGSKTVSQCKNFYFNKKRONLDELQOHLKMEKERNARRKKKAPAAASE 692
Db 244 DESP-SRM-----RTRN-----KQNNNSTSSGNVTAGN 272
QY 693 EAAFPVVEDEEMEASGVSGNEEEMVVEAEALHASGNEVPRG----ECSGPATVNNSSDT 748
Db 273 GCGNATSISSGTTGGGAAGSSSSKQDSANAV-ANGKRPKRGSSETPDVSGASVDS----327
QY 749 ESIPSPHTAAKDTGONGPKPATLGADGPPGPPPTPPRTPSRAPTEPTPASEATGAPT 808
```

D	b	328	- - -PKTPTKAVASSANKRK- - - -GGQDTPNKKKTQEBSNEPS-AHEBNALKEK 375
Q	y	809	PPAPPSP- - -SAPPVVYPK-EKBEESTAAAPPVEEGEEKPPAAABELADVDTGKAEE 860
D	b	376	RKRPDSPVESMNSDSRSDSVLDDGESNTTDTTA- - -EQGSTKDSKE-TWSCKEERE 428
Q	y	861	PVKSECTEAEAGPAKGDABA-AEATAAGALKAEKKEGSGRATTAKSGAPODSS- 918
D	b	429	MVTNDLEAKAEE--KAIEALAEDSKSAIKNMDEE- - - -TNIQAPSSADTSL 476
Q	y	919	- - -ATCSA--DEVDEAEGDGKNLLSPRSLLTPTGDP 952
D	b	477	V DGNPNALPSVPAIPTMKVPTIATVEALNASVDREAKEIEMESCDSDPEMLKKLATIK 536
Q	y	953	ANASPOKPLDLKQLKORAAAIPPIQTIVKVHPPREDAAPTK-PAPPAPPPPQN- - - 1004
D	b	537	QEVSPQQOQHWOQSQ- - - -QMOQOLAPVGIPOPPSCPSESUYIKKEP 582
Q	y	1005	LQESPDAQPGSGSPROKSPPADPADEAFAPAQAOKLPDGPWCWTSGLPPFPVPREVIK 1064
D	b	583	MEDSMDATCNONGNEPQDLVKVIEIKNEDALKHSAGGLPPSGPC- - - - 626
Q	y	1065	ASPHPDPSAFSAPPG-HPLPLGLHDHTARPVLPRPTTISNPPLISSAKHSVLERQI 1122
D	b	627	- - -APPSALHPL- - - -SGAP--VESQCEPLHQ- - - 650
Q	y	1123	GAISOQMGSVOLHVPSY-SHAKAPGVPTMGLPLPMDPKLAPFSGVGKQEOLSRGOAGPP 1181
D	b	651	- - -HMPHGQVTTQPPGYLDG- - - -OLKYGSGQGCVPP 682
Q	y	1182	ESLGVPSTAQASVLRGTALGSVPGSITKGIPST--RVPSDSAITYRGSI THGTPADVLY 1239
D	b	683	- - -QPQLHSDAAGVSQA--PPCAPITTPQKYPPEMEMKF- - - -APQDLKY 724
Q	y	1240	KGITRIIGEDSPRLDRGEDSLPKGHVIYEGKHGHLVSYEGMSVTOCKSEDGRSSG 1299
D	b	725	- - -PPPPPLD- - - -ALKYSORMQAAAAA- - - 745
Q	y	1300	PPHETAAPKRTYM- - -MEGRVGRAISSATEGLMGRAIPER- - - 1339
D	b	746	- - -AAAAGKIDMKYMWEOGKYNVELSAAH- - - -QPSKPYQDSLKIPIDKPG 792
Q	y	1340	--HSPHLKEQHHRIGSITOGIPRSYVEAOEDYLREAKLLKREGTPPPPPSRDLTEAY 1397
D	b	793	FGLHPHNV- - - -GSP--LDAAHKY- - - -GPPTSQESQQQ 822
Q	y	1398	KTQALGPLKLPKPAHEGLVATVKAGRSIHIIPRELHRTPELPIARRP- - - -LKEG- 1449
D	b	823	QPO- - - -PPAQVPPGATPPPGAIA--MPKPHYQHDVQTPTPLGRPFEPFTGLMLKYGD 872
Q	y	1450	--SITQCTP--LKYPD--TCASTGSKKHVSRSLIGSPGTRTPPVHPLDVMDARALER 1501
D	b	873	PLAAKYGPQDLKYPMPPVQAGPADVKPYGGENLIKSSPYGPPSPSIDA- - - 923
Q	y	1502	ACYEESLKSRPGTASSSGSIARGAPVIVPELKGPROSPILTYEDHGAPFAGHLPRG- - - 1557
D	b	924	- - -SARSIFG-QDSQGSNSNSQPSMPBPQQQFQSPHPSHPMSPAGGLPGCMHPQ 976
Q	y	1558	- - -SPVTMRBPT- - - -PRLQSGLSUSSKASQDRKLTSTPRE 1591
D	b	977	NLIHGPPPPAAGSGGPOPPPPPTSLHOPTPTSAGPPSLQHLGFHGHQSLSVASSTPPS 1036
Q	y	1592	IAXSPH--STVPEHHHP-I SPVEHLLARGVGDVLYRESHIPLAFDPTSIPRGIPLDAAAA 1648
D	b	1037	SIGIPPTLSWASHMHPLHPLHAH-LQGL- - - -HRP- - - - 1068
Q	y	1649	YYLPRHLAPNTYPHLYPYLLIRGYPDTALENRQTIINDYITSQQMHNTATAMAQRAD 1708
D	b	1069	- - -HDLPSMHFAWPLSLQCHP- - - -QHCHGLPPSHTSQQOQQOQQOPGGPAG 1118
Q	y	1709	MLRGLSPRES- - - -SLALNYAAGPRGIIDLQVPH- - - -LPVL 1743
D	b	1119	TVAITPSAPOOPRRSMHDPQOSBRBPTTSOPSTTWAGSSGCG-PPPOOSPHAHRTSLPGL 1177

Qy	1744	V-----PPTPGT-----PATAMDLAVLPTA-----PQPFSSSHSSSPSLSPGG-----	1781
Db	1178	AGSGPPPGGLIGHPMATHPHLAHLPPCHPAHAALAHPGHLLSHSAGLPGCGPIALLA	1237
Qy	1782	-----PTHL-----TKPTTSSSERDRDRDRDREREKILTS	1817
Db	1238	GPGLGIPRESAISRRTPPHLPHSHASSAPLTAHVA-----SMTSTSMULT	1285
Qy	1818	TTTVEHAPIWR--PGTEQSSGSSSG--GGG--SSSRPASHAH-----	1858
Db	1286	TSTVPSAFGRSPSVQISSLSSGGSPGSPGVGPGMPNSSAAAAAHHRAASPASSVS	1345
Qy	1959	---QHSIPSPRTQDALCORSVLHNTGMKGIITAV-EPKPTVLTSTSTSSPVRPAATPP	1914
Db	1346	SLGRQSPHVPVQSPSLSHHPS--SSALSAASAAVAERDRHALMQOS-----PHMTTP	1396
Qy	1915	PATHCPL-GGTLDQGVTP-----TLMBPVLLPKAEPRVAR-PERPRADTGHAFIAKP	1963
Db	1397	PVSNASLWASPLSKMYAPOQQRGLTSPPHLRPGASPPVIRHPQWL-----	1445
Qy	1964	PARSGLEPASSPKSGSEPRLPVDPVSGHATIARTPAKNLAHPHASPPDPAPPASADPHR	2023
Db	1446	-----PLPLIAPGGGIPQIGVHPGQSPYPH-----PLLHPSVFYSPHH	1483
Qy	2024	EKTQSPFISQELSLRSLGVHGSSYSPGVPSVSSPSLTHDKGLPKHLELDKSHLE	2083
Db	1484	H-----PFN-----SPYGVAPYGP-----GFPAYMKP-----	1505
Qy	2084	GELRPKQPGVKLGG-EAHLPHLRPL-----ESQFSSSPLLQTAPGVKHQHVVTLAQ	2137
Db	1506	-----PQPGQLDPAAVMAAHAGLQGPFPQPMQRDEQNAASAAAAAQAAEKHQHAAAAA	1561
Qy	2138	HISEVITODY-----TRHHPOLSNAPLAPLYSPGASCVPVLDLRRPSDL----	2183
Db	1562	QHKAPQOQPGGMNPKPTPTKTPQGGGMP-----PGMGGPTPTGLPPGAVPGSH	1615
Qy	2184	---YLPPPDHG---APARGSPHS-----EGKRSPEPNKTSVLGCGEDGIEPV-----	2225
Db	1616	MPGYQPQPHGSPFAPQDQPHGLKPTSHMDALRAHASANSAGMGGGHHTPEFLIDIE	1675
Qy	2226	-----SP-----PEGM-----TEPGHRSVAVPLYLRDGEQTEFSRGMGSKPGNT	2265
Db	1676	PDPEETPSPTNIPRGPSPEAKPDDTECHRSQSALFVRHIDRGDYNSTCR-----	1726
Qy	2266	SQPPAFSKLTESNAMYKSKQEIKNKLNTNRNPEYNI SQPGTEIFNMPAITGTGLM	2325
Db	1727	--TDLIFKPVADSKLA---RKREDRDKLAERERRRQOQOQO-----	1764
Qy	2326	TYRSQAVQEHASTNMGLEAIIRKALMGKYDQWEESSPLSAN-----AFNPL	2371
Db	1765	--QOQOQOQAAAAQAAQAAKMA-----ELKPPYADTPALRQLSBYARPHVAFSPV	1814
Qy	2372	-----NASASL-----PAAMPITAADGR	2389
Db	1915	EQWVPYHHMGPMYRERELEEIKNAQAAAAASQSRDPHWMYYRGRIGHSPQPLIYANPAI	1874
Qy	2390	SDHTLTSPGGGGKAKVSGRPPSSRKAKSPAPGLASGDRPPSV	2430
Db	1875	SOMERERLIGTPPHHVGLDPGEHMVRMPPQPEAGFOLFPNV	1915

RESULT 19

RESULT: 1:
ABU08487

ABU0848 /
ID ABU08487 standard: protein: 8991 AA.

XX
ID
ABU06467

XX AC ABU08487:

AC ABU06467; XX

AA	24-JUN-2003	(first entry)
DT		

DI 00N-2003 (1115C enclty)
XX

DE *S. pneumoniae* pneumococcal surface protein A (PspA) protein.

[illegible]

KW Pneumococcal surface protein C; PspC; pneumococcal surface protein A;

[illegible]

RESULT 22	
AAU33195	AAU33195 standard; protein; 2153 AA.
XX	
XX	
XX	AAU33195;
XX	
XX	
DT	18-DEC-2001 (first entry)
XX	
XX	Novel human secreted protein #3686.
XX	
KW	Human; vaccination; gene therapy; nutritional supplement;
KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX	
OS	Homo sapiens.
XX	
FN	WO200179449-A2.
XX	
PD	25-OCT-2001.
XX	
PF	16-APR-2001; 2001WO-US008656.
XX	
PR	18-APR-2000; 2000US-00552929.
PR	26-JAN-2001; 2001US-00770160.
XX	
PA	(HYSE-) HYSEO INC.

xx Nucleic acids encoding a range of human polypeptides, useful in genetic
pt vaccination, testing and therapy.

XX
PS
Claim 20: page 727-728: 765pp: English:

The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon

CC and/or nerve tissue growth or regeneration; immune suppression and/or
stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
secreted proteins of the invention

XX CC

XX Query Match 4.0%; Score 522.5; DB 4; Length 2153;
Best Local Similarity 21.0%; Pred. No. 3.9e-20;
Matches 488; Conservative 213; Mismatches 809; Indels 813; Gaps 114;

XX Sequence 2153 AA;

QY 474 KKNENYKSLVRSYRRGKSGQQQQQQQQQ---QQQQQQQPMRPSSEEEKDEKEKEAE 530
DB 70 KGNDPNLSLVPKDGTGWASKQEOSDPKSKSDASTAQPPESQLPAS-----QTAS 119
QY 531 KEEBKPEVENDKEDLLKEKTDDTSGEDNDEKAIVASKGRKTANSQGRRKKGRITSRMANE- 589
DB 120 NQPKRPAAENTPLV-----PSGV--KSWAQSIVTHGAHGDDGRASSILLSRFSEEPF 170
QY 590 -----ANSEBAITPQOASBLA-----SMELNESSRWTE-----ERMETAKKLLEHGR 632
DB 171 PTLQAAGDQDKAAKERESAQSOGSPGLPQNSTTWRDGGRGPDBLE-GPDSKLHHGH 229
QY 633 N-----NYKKQNLDLILQHKMKWEKERNARRKKKAPAAASEAAPPV 699
DB 230 DPRGLOFGSPPQPFPYRGMMPPFYPLFPFPYPGQGYRYPTPDGSRFRPVAGPR 289
QY 645 TVSOCKNFYF-----NYKKQNLDLILQHKMKWEKERNARRKKKAPAAASEAAPPV 699
DB 290 GSGPMLRVFVGVRPSILKDNLKEFDQ-----LDQEND-----DGWAGAHEVDYTEK 338
QY 700 VE--DEEMEAHSVGSNEEEVMVEAEALHASGNEVPRGECSPATVNNSDTESIPSPHTE 757
DB 339 LKPSDEE--DGRSDDEGAEGHRDSQSASGERP-PEADGKK-GNSPSEP-PTPKTA 391
QY 758 AAK-----DTQNGPKPATLG-----ADGPPGPPTTP-----RTSRAPTE-- 795
DB 392 WAETSRRPETEGPPAPKPPLPGDYDRGGPCPKPAPEDEDEAMQRKQSSSETLSLA 451
QY 796 -----PTPASEATGAPTPPPAPPSPSPAPPVP 821
DB 452 VERARRREBERMQERRAACAEKULRLDFEGADPKLKAEPAAPPAAPTAPPPA 511
QY 822 VPKEKEETEAAAPVBEGBEQPPAAABELAVDTGKAEEVPKSECTEABEGPAKGDAE 881
DB 512 VPKE-----LPAPPA-----PP-----PASAPTPETEPEEPAQAAPQA 544
QY 882 AAEATAEGALKAEKKEGCGRAATTAKSSGAPQSDSSATCSADEVD-----EAEGGDKNRL 937
DB 545 STPTFT-----GVAAAPTIVSGGGSTGSTSGSPFEAGFVPEQLPSKGEPEEP 592
QY 938 LSRRPS-----LLTPTGD-----PRANASQPKPLDLKLQKQ-- 968
DB 593 VPPPTTPVKVEPKBGDIGTRPPSQGLGYPKYKXSLPFRQROQOEQL-LKQQOQH 651
QY 969 ----RAAAIPIQVTKVHEPREDAATKPAAPPAPPQNLQPEDSAPOFGSSPRGKS 1024
DB 652 WQHQCGSAPTPVPP-SPPQPVTLGAVPAPAAPPF-----PKALYFALGRPP----- 700
QY 1025 SPAPPAKEAF-----AAEAQKLPGDPPC--WTSG-L-PFPVPPRE-----VIKASP-- 1067
DB 701 -PMPDMNFDPRMWMIPPYVDVPRLLIQGPPPLEFFYPGVHPGSLVPRERSDSLGSSEPFDR 759
QY 1068 HAPDFSAYSAPGHPLPLGLHD--TARVPLPR-----PTISNP 1105
DB 760 HAPAWLERGTTPVDPKLAWGVFTATPAEPRLTSLRQAODEDDKGRSETPPVPPP 819
QY 1106 PPLISSAKHPSVLROICAISQGMVOLHYVPYSEHAQAPGVPTMGLPLP----- 1155
DB 820 PPYLAS--YGFPPN--GAPCPPLS--RPLEEPGPRP-----LPWPFGSDDEVAKIQ 865
QY 1156 -MDPKKLAPSGVTKQE--QL-----SPRCQAQPPESLGVPTTAQEAHVLRGTALGS 1202

Db 866 TPTPKKEPP-----KETAQLTGDEAGKRLPASRGAGPP-----PPRESRT--ETRWGP 914
 QY 1203 VPGSITKIGPSTRVPSDSAITVR--GSI-----THGTADVLTKGTITRIIGEDSP 1252
 Db 915 RFGSS--RRGP---PEFCAPRRAGPIKPPPTKVEELPKPLEQGDETPKPKPDP 969
 QY 1253 SRLDRG---EDSLPKGHVIEGKKHVLSE-----1281
 Db 970 LKTKKLGKPKETPPNGNLSAPARLDYSYERVGFTSCRGGRGEYFARGGRFGRTYG 1029
 QY 1282 ----GCMVYQCSKE---DGR--SSGPPHETAAPKRTYDMMEGVRAISSASIEGLWG 1332
 Db 1030 GRCGRGRSEFRSYREFRGGDGGGTGGNHPAPR-----GRHASETRSEGEY 1079
 QY 1333 RAIP-----PERH---SPHLKEQHIRGSIQO-----GIPRSYVEAQED 1369
 Db 1080 BEIPKRCRQSGTSGTHESDLAPSKAPTKEGTLAQVPLAPPPGAPSPAPART 1139
 QY 1370 YLRREAKLLKREGTP-----PPP-----PPSRDLTEAYKTAQLGLPKLKAHEG 1413
 Db 1140 -CRGGRVFTPRGVSRRRGRGGRRPPQVCPGMSPPAKSL--APKKPTGP--LPSKEP 1194
 QY 1414 LVATVKEAGRSIHEIPRELHHTPELPLA-----PRPKESITQGT 1456
 Db 1195 L-----KEKLIPGLSFVARGSGNGSNVGMEDGERPRRRRHGRAQOQK 1239
 QY 1457 ----LKVDGTASTGSKKHVRSILIGSPRTFP-----PVHPLDWMADARALERAC-- 1503
 Db 1240 PPRFRLKQERENARGSE-----GKPSLTLPASAPGEALITVIVAPPPAAAK 1291
 QY 1504 ----YEESLKSRPGTASSGSGSIARGAPVIVPE-LGKP-----1536
 Db 1292 SPDLNSQNSQANEWEETASESDFTSERRGDXEAPPPVLLTKAVGTGGGGGAVGI 1351
 QY 1537 ----RQSPLYEDHGAFFAGHLPRGSPVTWRETPRLQESLSKAS-----1580
 Db 1352 SAMSRLDQORAKDLKRS-----FSSQRP--GMEQRNRRPGGKAGSSGSSGGGGGPG 1406
 QY 1581 ----QDKLTSTPREIAKSPHST--VPEHHPHPISPYEHLRGV-----SGVDLYR 1625
 Db 1407 GRTGPGRGKRWSPKNSRPPPEERPPGLPLPPPPSSAVFRLOQVTHSNPAGIQQAL 1466
 QY 1626 SHIPLAFDPTSIPIRGIPLOAAAAIYFRLHAPNTYPHYLYPYLIRGYPDTAALNRQTI 1685
 Db 1467 AOLSSRQSGSVTAAGGHRHKPGPPQAPQ--GPSPRPTRYPEQR-----1508
 QY 1686 INDYITSQQMHNATAMQAQMDMLRGL--SPRESSLALNYAAGPRGIDLSQVPHLPVL 1743
 Db 1509 VNSGLSSD-----PHFEPGPMVRGVGTGTPRDSAGVSPFPKRR-----ERPPRKPEL 1556
 QY 1744 V-----PPTCGTGA-----TAMDLAVLPTAPO 1767
 Db 1557 LQESLPPPHSSGFLGSKPGPGQPAESRDTGTTEALTPHWNRLHTATSKSRTPRTSMEP 1616
 QY 1768 FSSRHSSPLSPGCPHTLTKPTTTSSEERDRDRDREREKSIILSTTTTVEHAPIW 1827
 Db 1617 W-----MEPLSP-----FEDVAGTENSQSDGVD-----LSGDSQVSSGPC- 1652
 QY 1828 RPTQESGSSGSSGGGGSSSPASHAHQHSPISTPTQDALQORPSVLHNTGMKGII 1887
 Db 1653 ---SORSSPDGGLKGAAGPKRPGGSSPLNAVPCGPGSEPPRPPAPHD--GDRKEL 1708
 QY 1888 TAVEPSKPTVL-----RSTSTS-----SPVRPA-ATFPPA-----TH 1918
 Db 1709 PREQPLPFGPIGTERSQRTDRGTEPGTIRSHRPGPPVQGTSDKSDLLVVGDSLKAE 1768
 QY 1919 CPLGGTLDGYVPTLMPEVLLPKBAPRVARPERPRADTGHAFKAPKPARSGLE-----1970
 Db 1769 KELTASVTEAIPVSRDOWELLPSAAS-AEPQKNLDSGHC--VPERSSSGQRLYPEVFG 1825
 QY 1971 ---PASPSKSGSEPRPLVPVSGHATARTAKNLAPHAS-----PDPPAPPASADPHR 2023
 Db 1826 SAGFSSSQISGGAMDLSLHNSGG---FRPGTPLSHYRQPLYLPFGPAPPALLSGVA 1882

Query Match 3.9%; Score 519; DB 4; Length 3536;

QY 2024 EKTQSKPFS-IQELRLSLGYHSGSSYSPGVPVSPVSPSLTHDKGLPKHLEELDKSHL 2082
 Db 1883 LKQQLDFSTMQATGELKLPAGGVLYPPPSF-LYSPAFCPSPLPDTSLLQVRDL-PSFS 1940
 QY 2083 EGELRPKQPCPVKLGEAAHLPLRP-----LPESQSSSP--L 2119
 Db 1941 DFYSTPLQ-----GGQSGFLSPGAPAOQMLLPVMSQLPVNVFGLPPAPPPLSL 1995
 QY 2120 LQTAFCVKGQRVWVTLAQHISEVITQDYTRHHP-----QQLSAPL- 2159
 Db 1996 LPVGAALQPPSFVVRPQSSPSTGVLPXLARPFPVYFGRTELHPVNKPFDFOKLSSNLG 2055
 QY 2160 -----PAPLYSPGASCVDLDR-----RPPSDLY--LPPPD- 2189
 Db 2056 GPGSSRTPTGRRPSSRLRSFGLNSRLQSLNSLTSGVFRNQASTFYQAGLPHPDALR 2115
 QY 2190 -----HCAPARGSPHSEGGKRSPEPNKTSVLGGGEGIEP 2224
 Db 2116 WIPKPWERTGRPRDGP-----SRAEPPGSR--GDKFGLPP 2151

RESULT 23

ABB65480
 ID ABB65480 standard; protein; 3536 AA.

XX ABB65480;

XX AC AC

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 23232.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09583.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 23232; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at fcp.wipo.int/pub/published_pct_sequences

XX Sequence 3536 AA;

Best Local Similarity 19.0%; Pred. No. 1.2e-19; Matches 589; Conservative 364; Mismatches 1037; Indels 1112; Gaps 136;		
QY	42 YOHSHRDYASHLSPGSIOPORRRPSELLSEFOP-----GNERSO-ELHLRPESHVYLP 95	1076 APNVSVHETE-PASPTPSKQL--MPTPTTPAPATKSPAA-----AAVAPFFCAAN 1124
Db	199 FOHLSQHADKY---FACQPCRLR-----FOADHLLGHQOQHSLHPEDGHATKPL 249	842 EQKPP-----AAELAVDTGKAERPVKSECT-----EAEPEGPAKDAABAEAT 886
QY	96 -GKSE-----MEFIESKRP-----RLLELLDPLLRPSPLLATGQFAGSEDLTKDRSLT 142	1125 RQTTNSTVTKSRKRLSDCIAMLTGKLEKLTQTEVPPGPGLOKEKEKEQQDPAKQDKT 1184
Db	250 MGADEPVLFRGLGTQNLRLSHRNQRKEEPPPLATPS-----SKSRSSS 295	887 AE-----GALKAKEGSGRATTAKSGAPO-----DSOSSATCSADEVDEAGGDKN 935
QY	143 GKLEVPSPSPHTDELELVPRLSKBELIQNMDRVREITWVEQOISKLKKKQOQLEE 202	1185 PQVDHQSPAAAVAEKER---VQPKTKRKAVERSRIIKMDATPETLAEPQER----- 1235
Db	296 -----RAATORQOQOHOHQOSQSHQO 317	936 RLLSPRPSLLTPTGDPRA-----NASQPKPLDLKQLKQRAAAIPIQVTKVHEPPREDA 989
QY	203 EAAKPPPEP-EKPVSPPPPI-----ESKHSRLVQIYYDENRKK---AEAHRILEGLG 249	1236 -----KPVAMPAPVPPVAPAPAILNAAPALQPPPIFASMEVAVVVPVAPVPPPVAAA 1290
Db	318 TAAAGNMALOESVNDPAVFRPLPASMCAVSSVASVDFDKFYKDVINVRHNLQNHLD 377	990 APTK-----PAP-----PQNLQSPESDAPQOQSSPRGSRSPAPADKE 1033
QY	250 PQV-----ELPLYNQPSDT---RQYHENIKINQAMRKKLIL--- 282	1291 APLVAAIVLEVAAPAPPPVATALVHPPTTRPTTKAAAKQMGAPP-----PKPPASLA 1344
Db	378 GRLFTSSSTTSATSSASHKSDLIMLHGPSQLPGGASYPTLLTADQFGGTGELLPLA 437	1034 AFABAQKLPDPPPCWTSGLPPPPPREVIKASPHA-----PPPSAFSYAPPGHP 1083
QY	283 YKGRNHAKQWKQFCQRYDQLEMALEKKVERIENPRRAKESKVRREYKQPPPEIRK 342	1345 A-----LKQYPPLEAT---LPVPTNSAPPVVAVPLVPVPPVPPVAMPPLGMQGW 1394
Db	438 KFRRRPHTKHSWKW-----DYVKKFTLINE 464	1084 LPLGLHDTARFVLP-----RPPTISNPPP-----LISSAKHPSVLERQIGALSO--- 1127
QY	343 QRELQERMO--SRVGORG-SGLSM---SAARSEHEVSEIIDGLSBOENLEKO--MRQAV 394	1395 VEVNYN--MLPKPLPLYMTQLQAPANPNPNPNPAPTAMFVMDHQPLNLTSGRGVLQKPLI 1452
Db	465 GGRLVKLLQPNYGLURDLSKJDMWTQLTWRQKHDLJL---SLTEQOQOERRLLEQLNL 521	1128 -GMSVQLHVPYSEHAKAPVGPV-----TWGLPLPMDPKKLAPFSGVKQEQOLSP 1174
QY	395 I-----PMLYDADQOR-IKFTNM---NGLMADP-----MKVYKDRQVMNMW----- 432	1453 SGTGDPGLVGLQHRAGMPARRQTCGFERNLIGLDMEMEPDLDSKSKSRKSPVPP 1512
Db	522 ILDHRLPHIILEQNEQAVIKCENBEDEGLDHPGDDLPTQTFADESFLSLLQLQPRGG 581	1175 --RGQAGPESLGVPVTAQEAASVLRGTALGSPVG-----SITKGIPS 1214
QY	433 -SEQEKETREKPMQ-----HPKNFGLI---ASFLEKKTVAE----- 465	1513 PRMQQELPMLVPLPLVTNANAL---APQSGGAPLPTFPVGGAPVTCQLPMSGVSGV 1569
Db	582 SSEQERQERDQERSRQKAVVLSGEWARPRLYLICCGAKFDQRKLEEHKTFRHSI 641	1215 TRVPSDSAITRGSTHGTTPADVLKGTITRIIGEDSPSRDLRGREDSLPKGHVIEGKK 1274
QY	466 CVLYLYLTKNENYKSLVRSY---RRGK-----SQOQOQOQOQOQ 504	1570 --VPAPLASHYYSNLDLILKIPOVRNPQGTGV-VVPSSGPPAVSASAAVAPHPVKSSG-- 1624
Db	642 YATHYEVVRELLAGNLLRHLFIPKALGRFAAASCIRWQIAPATVQPKQBOVQPE 701	1275 GHVLSYEGGMSVTQCSKEDGRSSGPPHETAAKPTDYDMEGRGVRAISASIEGLMGRA 1334
QY	505 QO-----QOQPMPRS----- 515	1625 GKRTTEG-----TGSKKD---NGKQAKACP-RMDEVNHHIDDALNSVIM--AVQAS 1673
Db	702 EETRASSASSRSEVSTPTPLSALEQOQSPASGSGNASSSSSCSPSSASTLLSTATR 761	1335 IPPERHSHPHLKEQBHHIRGSITQGIPIRSYVEAQEDYLREAKLLKREGTPPPPPSRDIT 1394
QY	516 -----SQUEK 520	1674 PPDDDEERQAKKE-----KEREREREREREKELEEPPLKPSNCI- 1714
Db	762 TSMNSTCTSSAATTVTSSTAPSSCTKGRKCSGLMDLYRHMLDSCSDYVWSLAKKK 821	1395 EAYKTQALGPLKLKAHEGLVATVKEAGRSIHEIPRELRHTPELP--LAPPLKEGSGIT 1452
QY	521 -----DEKEKEKEKEKEE-KPEVENDKEDLLKEKTDOTSGEDNDEKEAVAS 566	1715 -----MPLQCAEVLPSAVLEKSQAVNAIPAIVAPVAVVVPVVLAPQVPPVKNL- 1763
Db	822 YRYCCTKRRAFSKLQLSARKKEKLEVEEETGDDGDAEVDGDGEAEABEGEESVS 881	1453 QCTPLKYDTGASTTCKKHDRSLIGS-----PGRTFPPVHPLDMADARALERACYES 1507
QY	567 KGRKTANSQGRKRG---RITRSMAN-----EANGEEAITPOQSAELASMELN 610	1764 --TPKK-----RSMRSRTIDCSALLALEETLPAASLFCQICPLPVNGDAK-----EE 1808
Db	882 SSKLKNRPRQPSDAESIRKLENLPAKRI CKKIPVDNKA KAKAKAKAKAKSMAKSLV 941	1508 LKSREGTSSSGGSIGARGPVIV-PELGKPRQSPLYEDHGAPFAGHLPRGSFVTMREPT 1566
QY	611 ESSRWTEEMETAKGLLEH-----GRNWSAIARMVGSKTSQCKNFYFNKQKQNLDE 664	1809 MBSLP-----DPIAPTAVVEFOL-----PVQVQSGT 1836
Db	942 KSKSKSKQORSSTKRIYNGHLLRTTRSRGRSSVATGSSAAAAVAAQOQSRKQKQOQK 1001	1567 PRLOGSSLSSSKASQDRKLTSTPREIAKSPHSTVEHHHPHPISPYEHLLLRGVSGVDLIRS 1626
QY	665 ILQOHLKWEKERNARRKKKAPAAASEEAPPPVVEDEMEASGVSGNEEEMVEAEAL 724	1837 TSLRDKKTETVMPCTNSTTIPTATPLAESPE-----PDVP-----V 1874
Db	1002 QOQOQOQOQKQKLEKVKGPPEP-----PAAVESDQMPA----- 1038	1627 HIPLAFDPTSTIPRGIPLD-----AAAAYLPRHLAPNPTVPHLYPPVIRGYPDTAALNR 1682
QY	725 HASGNEVPREGCSGPA---TVNNSSTESIPSPHTEAANDTQONGPKPPATLCADGPPPG 781	1875 NLPOA--FTSVAPVPPVSVISVAPAVLPLPTRTATPP-----PTTMAETNC 1918
Db	1039 -----KTEAPAPAPTVAKSRSFAEEKPSLRLEL-----PQTL-----PEAA 1075	1683 QTIINDYITSOQMHHNTATAMAQADMLRGLSPRESSIALNYAAGPRGIILDSQVHPLV 1742
QY	782 PTPPPRTSRAPLEPTPASEATGAPTPPPAPSPSPAPPPVVPVKEEKEETEATAAPPVEGE 841	1919 SSLMEEH--SSNLNNNTSS-----GPHSLAQ-SEQPI 1947
		1743 LVPPTFGTPTAMDRLAYLPTAPOFSSSRHSSPLSPGPGPHLTXTPTTSSSERERDRDR 1802
		1948 PTAATTABEAPPVKEDEELPAKKK--QRRRRKNELAAIVAOQLLESFKIDNARRNLKJL 2005


```
Db 2558 SEVLNMQSEYRLHPYTPVRDVRIMVHPHTAVSEQPR-----ADGWVKVPBASKA 2608
QY 2232 TEPGHSRAVYPLLYLDEGEOTEPSPRGMKSPGNTSQP 2268
Db 2609 PQ-----QPGKEAAKTPDAAKAP 2626

RESULT 25
ABB33490
ID ABB33490 standard; peptide; 2665 AA.
XX ABB33490;
AC ABB33490;
DT 04-FEB-2002 (first entry)
XX
DE Peptide #996 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
XX
OS WO200157277-A2.
XX
PN 09-AUG-2001.
XX
PD 30-JAN-2001; 2001WO-US000669.
XX
PF 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human foetal liver.
XX
XX Claim 27; SEQ ID NO 26125; 639pp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: the sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2665 AA;
XX
Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

QY 71 EFQPGNERSQELHLPF-----SHSYLPGLGKS-----EMFETESKR---PRLELLP 114
Db 525 ERKSGQKSHSVNTEBKIGIDIDHTOSYRQMEQSRKQKQMEIAKSEKSPKXDV-- 582
QY 115 DPLLRSPLLATQCPAGSDELTKDRSLTGKLEVPSPSPPTDPELELVPPRLSKEELIQ 174
Db 583 DEYERSLVHEVGKP--PQDVTD-----SPFSKK-----X 611
QY 175 NMDRVDEI--TWVQOISLKKKKQOOLBEAAKPP-----EPEKVPSPPIES----- 221
Db 612 RMDHVDICTKERNYRSRQISEDSERTGGSPSVRHGSFHEDEPIGSPRLLSVKGSP 671
```

```
QY 222 -----KHSRLVQIYYDENRKK--AEAHRILEGLGPOVELPLYNQPSDTRQ 265
Db 672 KYDEKVLPSYNTVREESLKFNPYDSRRQEMADMAKIKLSVLNSEDENRM-----DSOM 727
QY 266 YHENIKINOAMRKKLILYFKRRNHARKQWKFCQBYDQIMEALEKKVERIENPPR--- 322
Db 728 KQDAGRFVSPNSII-----KRDLSRKRSVRDL--EPGEVPSDSDESDHSHKSPASAL 781
QY 323 -----RAKESKVREYVEKQPEIRKQRELQERMOSQVQQRGSGI-----SMSAAR 367
Db 782 YESSRLSFLLRDEYKLRERDERLSSLERNFYSFALDKTITPTKALLERAKSLSSSR 841
QY 368 SEHEVSEIIDGLSE-----QENLEKQMRQLAVIPPLYDADQQRIKFINNGLMADPMKY 423
Db 842 EEN--WSFLDWDSTRFANFRNNKDEKVDSPRIPSWYMKKKIR--TDSEGMDDKDEH 898
QY 424 KQOVNMWSEBOKETFREKFMQHPKNFGLIASFLERKTVAECVLYYYITKNENYKSLV 483
Db 899 KEE-----QERQELFASRFL-HSIFQDQSKRLQ-----HLERKEEDSDFIS 940
QY 484 RSYRRRGKSGQQQQQQQQQQQQQQQ-----MPSROBEKDEKEKEKEKEE 533
Db 941 GRIY---GK---QTSSEGANSTTDSIQEPVVLPHSRFMELTRMQCKEKEKQPKVEKOE 994
QY 534 EKPEVENDKEDLLKEKTDGTDGNDDEKAVASKGRKTANSQGRKRGRITRSMANEANSE 593
Db 995 ---DTEN-----HPKTPESAPEND-----SELKTP 1017
QY 594 EAITPQQAELASMEINNESSRWTEBEMETAKGLLEHGRNWSAIARMVSKTVSQCKNFY 653
Db 1018 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1046
QY 654 FNYKKRQNLDELQOHLKWEKERNARRKKKAPAAASEEAPPPVVEDEMEASGVSGN 713
Db 1047 -----EAPLVTEKTVETPATVSEAKPASPAPA 1075
QY 714 EEMVVEEAELHASGNVPRGCSGPATVNNSSDYESIPSPHTEAAKDTGQNGPKPATL 773
Db 1076 PVEQLEQV-----DLPPGA-----DPDKAANMP---AGVESSGSDQPP-YL 1114
QY 774 GADGPPGPPPTPRRTSRAP--IEPTPASEATGAPTPPPAPSPSPAPPVVPVKEKEEBET 831
Db 1115 DAK-----PPTGASFSQAESNVDPEDPS-----TQPLSKDAQKSEANEKPAEKPDAT 1163
QY 832 AAAPP-----VEEGEQRKPPAAEELAVD---TGKAEHPVKSECTEBAEAGPAGKDAEAA 883
Db 1164 ADAEPDANOAKAAAPESQPPASEDLEVDPPVAAKOKKPNKSKRSTPVOAAAIVSEKPV 1223
QY 884 EATAEGALKAEKKEGGSGRATTAKSGAPQDSDS--SATCSADEVDEAEGDKNRLLSRPP 942
Db 1224 TRKSERIDREKLRNSPRGEAQKLELQWEAEKTRTASKNSAADLE-----HPBP 1275
QY 943 SL-LTPT-----GDPRANASPOK-PLD-----LKOLKQRAAIIPTQVTK 980
Db 1276 SLPLSRTRNRNRSVVATMGD--HENRSVKPEVQPRVTRKLERELQEAAPV---TTPR 1332
QY 981 VHEPPR-----EDAAPTKPAAPP-----PONL-----QPSDAP 1012
Db 1333 RGRPPKTRRRARDEBENEAKPAETLKPPEGWRSRPSQKTAAGGGPQGGKKNPEKVDAT 1392
QY 1013 QQPGSSPRG-----KSRSPAPPADKEAPAAEAQKLPDGPCCWTSGLPFPVPPREVIKASP 1067
Db 1393 REATEIVGQIVGVKESMEPKAAEEAGSEKQRKRD--AGTDKNPPTAPVETVVEKKP 1450
QY 1068 HAPDPSAFSYPAPPHPLPLGLHDTARPVLPFRPTTISNPPPLISSAKHPSVLERQIGAIQ 1127
Db 1451 -APEKNSKS-----KEGRSRNSRLAVDKSASLKN-----VDAAVSPRGAQAQERES 1497
QY 1128 GMSVOLHPVYSEHAKAPGVFTWGLPLMDP-----KKLAPFSG-----VKQBL--- 1172
Db 1498 GV-VAVSPKESSEPOKEDGLSQLKSDPVDPDKEPEKEDVSASGSPPEATOLAKOMELBO 1556
```

QY 1173 -----SRGQAGPPESLGVPTAQEASVLRGTALGSV----- 1203
Db 1557 AVEHIAKLAESAASAAKADAEGLA--PEDRDKPAHQASQETELAAAGISINDISGEPE 1614
QY 1204 -----PGSITKGIPTRV-----PSDSALTYRGSITHGTPADVLYKGTI--TRIIGED 1250
Db 1615 NFPAPPYPGESQTDLOPPAGAAQOPSE-----EGMETDEAVSGIILETAATES 1664
QY 1251 S-----PSRLDRGRDLSLPKHVIVYEGKKGHVSLEGGMSVTQCSKEGRSSS 1298
Db 1665 SRPPVNPAPDSAGPTDKARGNSSETSHVPEAKGSK-----EVEVTIVRKDKGRQ-- 1716
QY 1299 GPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPPERHSPHLKEQHIRG---SI 1355
Db 1717 -----KTRRSRKRNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
QY 1356 TQGISRSYVBAQEDYLRRRAKLLKREGTPPPPPPPSRDLTEAYKTQALGPLKPAHEGLV 1415
Db 1753 NEGTTVQHPEAFQ-----BEKQSEKPHSTPPQOSTSDLSKIPSTE----- 1792
QY 1416 ATVKEAGRSIHETPRELRHT-----PELPLAPRPLKEGSIQTGTPKLYDTCGASTTGSKK 1470
Db 1793 -----NSQELSVBERPTTKASVDPDLPDPQP-----APVDEPOA-----R 1830
QY 1471 HDVRSLLIGSPGRTPP--PVHPLDVMDARALERACYEESLKRPGTASSSGSIAAGAPV 1528
Db 1831 FRVHSIIESDPVTPPSDPSIPIPTLPSV-----TAAKLSPPVASG--- 1870
QY 1529 IYPELGKPRQSP-----LTVEDGAPFAGHLPRGSPVTMREPTPRLOEGSLSS---SK 1578
Db 1871 -----GIPHQSPTKVTEITRQE-----EPRAQSTPSPALPDPTKASDVDTFS 1913
QY 1579 ASQDKLSTPPEIAKS-----PHSTVPEHPHPISPVEHLLRGVSGVDLYR 1625
Db 1914 SSTLRKILMDPKVTSATSVTSTVTTAIEPVSAAPELHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDPTSIPIRGILDAAYLYPHRLAPNPTYPHLYPIRGYPDTAALENRTI 1685
Db 1968 KTAPPVTNNSEIQASEVLVAADKEKVAVPVIAKIT-----SVISRMPPVSIDLENSQKI 2020
QY 1686 INDYITSOQHHNTATAMAKQADMLRGLSPRESSLALNVAAGPRGIIDLQVPHLPVLVP 1745
Db 2021 -----TLAKPAPQTLTGL-----VSALTGLVNVSLVP--VNALKG 2053
QY 1746 PTPGTATAMDRLAYLPTAPQPPSSRRSSPLSPGGPTHLTPTTTSSSERERDRDRER 1805
Db 2054 PVKGSVTTLLKSLVS-----TPAGPVNVLGPV----- 2080
QY 1806 RDREREKSILTSTTTVEHAPI-----WRPTEQSSGS-----SGSS 1841
Db 2081 -----NVLGTGVNVLTTPVNATVGTVNAAPGTVNAASAASAVNATASAVTVTAGAVTAAS 2133
QY 1842 GGGGGS-----SSRPASHAHQHSIPSPRTQDALQORPSVLHNTGMGIIT 1888
Db 2134 GGVATTGTVTWAGAVIAPSTKCKQASANENSRRFPGSMPIVDRPA---DAG--SGAGL 2189
QY 1889 AVEPSKPTVLRSPS---TSSPVKPA---TFPPATHCPLGGTLDGVYTLMEVLLPKEA 1942
Db 2190 RVNTSEGVLLSVGGQKTGQPRISAKISQIPPAS-----AMDIEFQOSVSKSQVKPDS 2243
QY 1943 PRVARP--BRPRADTGHAFIA-----KPPARSGLEPASPSKSGSEPRPL--- 1984
Db 2244 VTASQPPSKGQAPAGYANVATHSLTLVTAQTNASPVISSVK--ADRPDL--EKPEFIHLS 2301
QY 1985 -----VPPVSGHATTARTPA----- 1999
Db 2302 VSTFVTGGTGVKVLTOGINTFPVLVFNQLVLTFSIVTTNKKLADPVLTKIETKVLQANL 2361
QY 2000 -KNLAPHASDPDPAPPASADP--HREKTSQKPF--SIQELERLSLGVHSGSYSPGVE 2054
Db 2362 GSTLTTHH---PPALPSPKLPTENVHVPSPGSPISPADRTVSHLAAAKLDHSPSPGPGPS 2417
QY 2055 PVSVPSPSLTHDKGL-----PKH---LBELDKSHL- 2082

Db 2418 SPFRASHFSSASTALSTNATVMLAAGIPVQFIISSIHPEOSVIMPPHSITQTVSLSHLS 2477
QY 2083 EGELRPKOPG---PVKLGGEAAHLPHLRPLPESQSSPSSPLLQTAGVKGHQVRVVTLAQH 2138
Db 2478 QGEVRMNTFTLPSIYYSIRPEALHSER--APL---OP-----QQIEVRA-- 2516
QY 2139 ISEVITQDYTRHHPOOLSAPLPAPLYSPFGASCPVLDLRRPESD---LYLPPPDHGAPAR 2195
Db 2517 -----PQRASTPQAP-----AGVPALASQHPPEERVHVHLPVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGEDGIEFPVSPPEGM 2231
Db 2558 SEVLVMQSEYRLHYPTVPRDVRIMVHPHTAVSEQPR-----AADGVVVKVPPASKA 2608
QY 2232 TEPGHSRSNAVVELLYRDGEQTEPSRMGSKSPGNTSOP 2268
Db 2609 PQ-----OPGKAAKTPDAKAP 2626

RESULT 26
AAM26950
ID AAM26950 standard; protein; 2665 AA.
XX AC AAM26950;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #987 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488997/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 27219; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP;
XX CC see AAI31315-AAI5746). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders
SQ Sequence 2665 AA;

Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

Db 2244 VTASQPPSKGQAPAGYANVATHSTLVLTAQTVNASPVISVK-ADRPRL-EXPEPIHUS 2301
QY 1985 -----VPPVSGHATARTPA----- 1999
Db 2302 VSTPVTQGGTVKVLTOGINTPPVLVHNQVLTPSIVTTNKKLADPVLTKIETKVLQPANL 2361
QY 2000 -KNLAPHASPPPPASADP--HREKTQSKPF--STQLELRLSLGVHGSYSPEGVE 2054
Db 2362 GSTLTTPHH-----PPALPSKLPTVEVHNVPSPGSPITPADRTVSHLAAAKLDAHSRPSGPGPS 2417
QY 2055 PVSVPSSPLTHDKGL-----PKH--LEELDKSHL- 2082
Db 2418 SFRASHPSSTASTALSTWNLVLAAGIPVPOFISIHPEQSVIMPPHSITQVLSLHUS 2477
QY 2083 EGELRPKQPG-----PVKLGEAAHLPHLRPLPESQSSPPLQATAPGVKGHQRVTVLAQH 2138
Db 2478 QGEVNMVTLPSITYSIRPEALHSR-APL---QP-----QQIEVRA-- 2516
QY 2139 ISEVITQDYTRHHPOQLSAPLPAPLYSFFGASCPVLDLRRPSPD---LYLPPPDHGAPAR 2195
Db 2517 -----PQASTPQAP-----AGVPALASQHPPEEEVHVHLPLVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGGEGDIEPVSPPGCM 2231
Db 2558 SEVLVMQSEYRLHPYTVPRDVRIMVHPHTAVSEQPR-----AADGVVVKVPPASKA 2608
QY 2232 TPGHRSVAVPLLYRDGSEOTEPSRMGSKSPGNTSOP 2268
Db 2609 PQ-----QPGKEAATPDAKAP 2626

RESULT 27
ID ABB28314
XX ABB28314 standard; peptide; 2665 AA.
AC ABB28314;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human peptide #965 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
PS Claim 27; SEQ ID NO 11282; 327pp + Sequence Listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2665 AA;

Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

QY 71 EFQPGNERSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR-----PRLLELP 114
Db 525 ERKSGQEKSHSVNTEEKIGIDIDHTQSYRKQMEQSRKQKQMEMETAKSEKFGSPKDV-- 582
QY 115 DPLLRPSPLATGQAGSEDLTKDRSLTKLEPVPSPPPHTDPELELVPRLSKLELIQ 174
Db 583 DEYERRSLVHEVGKP--PQDVTD-----SPPSKK-----K 611

QY 175 NMDRVDREI-TWVEQOISKLKKQOOLEEAAKPP-----EPEKVPSPPPPIES----- 221
Db 612 RMDHVDFDICTKERNYRSSRQISDSERTGSPSVRHGSHFDEDPICGSPLLSVKSGSP 671
QY 222 -----KHSRLVQIYDENRKK--AEAAHRIEGLGQVQLPVLNQPSTDTQ 265
Db 672 KVDEKVLPSYNTVREESLKFNPYDSRRQEMADMAKIKLSVLNSEDENRW-----DSOM 727
QY 266 YHENTIKINQAKKKLILYFKRNHARKQWKQFCQRYDQLMALAEKVKYRIENRR-- 322
Db 728 KQDAGRFDVSPFNSII-----KDSLRKRSVRDL--EPGEVPSDSDEDEGHHKSHSPRASAL 781

QY 323 -----RAKESKVEYVEKQPFIRKQRELQERQSRVQSGSGL-----SMSAAR 367
Db 782 YESSRLSFLLRDREKLRDERLSSSLERNKFYSFALDKTITPTKALLERAKSLSSSR 841
QY 368 SEHEVSEIIDGLSE-----QENLEKQMRQLAVTPPMLYDADQORIKFINNGLMADPMKVY 423
Db 842 EEN--WSFLDMSRPFANFRNNKDEKVDSPAPRIPSWYMKKKIR--TDSEGMDDKKEDH 898

QY 424 KDRQVMNMWSEQKETFREKMQHPKNECLIASFLERKTVAEVCVLYYLTKKNENYKSLV 483
Db 899 KEEE-----QERQLFASRFL-HSSIPEQDSKRLQ-----HLERKEEDSDFTS 940
QY 484 RRSYRRRGKSQQOQQOQQOQQOQQOQQOQ-----MPRSSQBEKDEKEKEKEAEKEE 533
Db 941 GRIY--GK--QTSEGANSTTDSIQEPVVLPHSRFPMELTRMQQKEKQDKQKPEVKEQE 994

QY 534 EKPEVENDKEDLLKXKTDGTDGDNDEKBAVASKGRKTANSQGRKRGITRSMANEANSE 593
Db 995 ---DTEN-----HPKTPESAPENKD-----SELKTP 1017

QY 594 EAITPQOSNELASMEINNESSRWTESEMETAKGELLEHGNWMAIARMVSKTQVSOCKNFY 653
Db 1018 PSVGPPSV-----TVVTLSSAP-----SALEKTTGDKTV----- 1046
QY 654 FNYKKRQNLDELQOHLKMEKERNARRKKKAPAAAGSEAAFPVVEDEEASGVSN 713
Db 1047 -----EAPLVTEKTVETPATVSEAKPASEPAPA 1075

QY 714 EEMVEEAEALHASGNEVPRGECGSPATVNNSSDTSIPSPHTEAAKDTGQNGKPPPATL 773

PD 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000666.
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 DR Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts.
 XX Claim 15; SEQ ID NO 20720; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 2665 AA;
 SQ
 Query Match 3.9%; Score 518; DB 4; Length 2665;
 Best Local Similarity 19.3%; Pred. No. 9.2e-20;
 Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;
 QY 71 EFQPGNERSQELHLRPE-----SHSYLPELGS-----EMEFIESKR---PRLELIP 114
 DB 525 ERKSGQKSHSVNTEBKIGIDHTQSYRKQMEQSRKQKQMEIAKSEKFGSPKDV-- 582
 QY 115 DPLLRSPLATGQAGSEDLTKDRSLTCKLEPVPSPPHDTPDELELVPRLSKEELIQ 174
 DB 583 DEYERSLVHEVGK--PQDVTDD-----SPSKK-----K 611
 QY 175 NMDRVDRDI-TWVEQOISKLKKQKQLEEEAAKPP-----EPEKVPSPPIES----- 221
 DB 612 RMDHVDFDICTRERNYRSRQISEDSERTGSPSVRHSFHEDEPIGSPRLLSVKGSP 671
 QY 222 -----KHSRLVQIYDENRK--AAAHRILEGPGQVELPYNQSPDTRQ 265
 DB 672 KVDEKVLPSYNTVRESLKFNPYDSRREQADMAKIKLSVLNSEDENRW-----DSQM 727
 QY 266 YHENIKINQAMRKLLILYFRNRHARKQWKFCQRYDOLMEALEKKVRIENPR--- 322
 DB 728 KQDAGRFVSPFNSII---KRDLSLRKSRVRL--EPGEVPSDSDEDEGKHSHSPRASAL 781
 QY 323 -----RAKESKVREYVEKOPPEIRKQRELQERMQSRVQQRGSL-----SMSAAR 367
 DB 782 YSSRLSFLLRREDKLRDERLSSLSLERNFYSPALDKTTTPDTKALLERAKLSLSR 841
 QY 368 SEHEVSEIIDGLSE-----QENLEKQMRQLAVIPPMYLDADQOQRIKFINNGLMADPMKY 423
 DB 842 EEN--WSFLDWDSRFANFRNNKDEKVDGAPRIPSWYMKKKIR-TDSEGRKMDKKEDEH 898
 QY 424 KQOVNMNMSQEKETFRFKFMOHPKNGFLIASFLERKTVASCVLYYYLTKNENYKSLV 483
 DB 899 KEE-----QERQLFASRFL-HSIFQDSKRLQ-----HLERKEEDSDFIS 940
 QY 484 RRSYRRRGKSQQQQQQQQQQQQQQQQQ-----MPRSSQEKDEKEKEKEKEE 533
 DB 941 GRIY---GK---QTSSEGANSTTDSIQEPVVLPHSRFEMELTRMOQKEKEDQRPKEKE 994
 QY 534 EKPEVENDKEDLLKKTDDTSGEDNDEKEAVASKGRKTANSQGRKGRITRSMANEANSE 593
 DB 995 ---DTEN---HPKTPESAPENKD-----SELKTP 1017
 QY 594 EAITQQSNELASMEINNESSRWTEEMETAKGLLEHGRNWSAIARMVGSKTVSQCKNPFY 653
 DB 1018 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1046
 QY 654 FNYKKQNLDELLOQHLKMEKERNARRKKKAPAAAGSEAAFPVPVVEDEEAEASGVS 713
 DB 1047 -----EAPLVTEKTVEPATVSEAKPASEPAPA 1075
 QY 714 EEMVVEEAALHASNEVPRGECSPATVNNSSDTPSPHTEAAKQTGONGPKPPATL 773
 DB 1076 PVEQLEQV-----DLPPGA-----DPDKAAAMP---AGVEEGSGDQDP-YL 1114
 QY 774 GADGPPPGPPTPPRKTSRAP--IEPTPASEATGAPTPPAPSPSPSPAPPVVPVKEKEEET 831
 DB 1115 DAK-----PPTPGASFSAESNVDPEDPS-----TQPLSKPAQKSEANEAPKRPD 1163
 QY 832 AAAPP-----VEEGEQKPPAAAEELAVD---TGKAEPPVKSECTEEAEEGPAKGDAEAA 883
 DB 1164 ADAEPDANKAAAPESQPPASEDLEVPVNAKKKPNKSKRKTPTVQAAAVSIVEKPV 1223
 QY 884 EATAEGALKAEEKGSGRATTAKSSGAPQSDS--SATCSADEVDEAEGDKNRLSPRP 942
 DB 1224 TRKSERIDREKLKRGNSPRGEAQKLELMEAEKTRTASKNSAADLE-----HPEP 1275
 QY 943 SL-LTPT-----GDPRANASQK-PLD-----LKQLKQEAALPPIQVTK 980
 DB 1276 SLPLSRTRRRNRSVYATMGD--HENRSPVKEPVEQPRVTRKRLERLOEAAAVP--TTPR 1332
 QY 981 VHEPPR-----BDAAPTTPAPPAPP-----PQNL-----QPESAP 1012
 DB 1333 RGPPTKTRRRADEEENEAKEPATLKPPEGWRSPRSQKTAAGGPGQKKGNEPKVDAT 1392
 QY 1013 QPGSSSPRG-----KSRSPAPPAKFAFAEAQKLPDPPCWTSGLPPLPPVPVPREVIKAP 1067
 DB 1393 RPEATTEVGPIQGVKSESSMEPKAAEEAEAGSEQKRD--AGTDKNPPTAPETAVEVVEK 1450
 QY 1068 HAPDPSAFSAYAPPCHPLPLGLHDTARVLPREPPTISNPPPLISSAKHPSVLEROI 1127
 DB 1451 -APEKNKS-----KRGSRNSRLAVDSASLKN-----VDAAVSPRGAAAQGERES 1497
 QY 1128 GMSVOLHVFPYSHAKAPVGTMGPLPMDP-----KKLAPFSG-----VKQEQ-- 1172
 DB 1498 GV-VAVSPEKSESPQKEDGLSOLKSDPVDPKPEKEDVSASGSPSPATQAKQWLEQ 1556
 QY 1173 -----SPRQAGPPESLGVTAEASVLRGTALGSV----- 1203
 DB 1557 AVEHIAKLAESAASAAKADAPGLA--PEDRDKPAHQASETELAAGIISIIINDISGE 1614
 QY 1204 -----PGSITKGIPTRV-----PSDSAITYRGSITHGTPADVLYKGTI--TRIIGD 1250
 DB 1615 NPPAPPYPGESQTDLOPPAGAQAQOPSE-----EGMETDEAVSGILLETAAATES 1664
 QY 1251 S-----PSRLDRGREDSLPKGHVIVEGKKGHVLSYEGGMSVTQCSKEDEGRSS 1298
 DB 1665 SRPPVNAVDPSAGPTDTKEARGNSSETSHSVEAKGSK-----EVEVTLVRKDKGRQ-- 1716
 QY 1299 GPHETAAPKRTYDMMGEVRGIRAISSASIEGLMGRAIPPERHSPHHLKQHHIRG-----SI 1355
 DB 1717 -----KTRSRKRKNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
 QY 1356 TQIGIPRSYVEAQEDYLRRREKLLKEGTPPPPPPPSRDLTEAYKTOALGPLKPAHEGLV 1415
 DB 1753 NEGTTVQHPAEPQ-----EKOSEKPHSTPPQOSTSDLSKIPSTB----- 1792


```
QY 1416 ATVKEAGRSIHETPRELHRT-----PELPLAPRPLKEGSIOTGTPLKYDVTGASTTGSKK 1470
Db 1793 -----NSSQEISVEBETPTKASVPPDLPPPPQ-----AFVDEBPQA-----R 1830
QY 1471 HDVRSLLIGSGRTRFP--PVHPLDVMADARALERACVEESLKSRRPGTASSSSGGSIARGAPV 1528
Db 1831 FRVHSIIEDPVTTPSDPSIPIPTLSV-----TAAKLSPPVASG--- 1870
QY 1529 IVPGLKPRQSP-----LTYBDHGAPFAGHLPRGSPVTMRPTPRLOEGSLSS---SK 1578
Db 1871 -----GPHQSPPTKVTEMITRQE-----EPRAQSTPSPALPDPTKASDVDTIS 1913
QY 1579 ASQDKLTSTPREIAKS-----PHSTVPEHHPIPISPVEHLLRGVSGVDLVR 1625
Db 1914 SSTLRKILMDPKVYSVTSTVTTAIAEPVSAAPCLHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDPTSI PRGIPLDAAAAYILPHRLAPNPTYPHLYPPYLIRGYDPDTAALENRQTI 1685
Db 1968 KTAAPPVTNNSEIQASEVLVNADIEKAVPVIAPKIT-----SVISRMFVSIIDLENSQKI 2020
QY 1686 INDYITTSQMHNTATAMAOQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745
Db 2021 -----TLAKPAPQTLTGL-----VSALTGLVNVSLVP-VNALKG 2053
QY 1746 PTPGTATAMDRLAYLPTAPQPFSSRHSSPLSPGGPTHLTKTPTTSSSERERDRDRD 1805
Db 2054 PVKGSVTTLLKSLVS-----TPAGPVNVILKGPV----- 2080
QY 1806 RDREREKSILTSTTVEHAPI-----WRPTEQSSGS-----SGSS 1841
Db 2081 -----NVLTGPNVNLTTVPVNAVGTVNAAPGTVNAASAVNATASAVTVTAGAVTAAS 2133
QY 1842 GGGGGS-----SSRPASHSHAHQSPISPTODALQOPRPSVLHNTGMKGIIIT 1888
Db 2134 GGVATGTGTVMAGAVIATSKCKQASANENSRFHGSPMVIDDRPA---DAG-SGAGL 2189
QY 1889 AVEPSKPTVLRSPS---TSGPVRAA---TFPPATHCPLGGTLGQVYPTLMEPVLPLPKGA 1942
Db 2190 RVNTSGVLLSYGQKTEGPQIRISAKISQIPPAAS-----AMDIEFQGSVSKSQVKPDS 2243
QY 1943 PRVARP--ERPRADTGHAFIA-----KPPARSGLEPASPSPKSGEPREL--- 1984
Db 2244 VTASQPPSGPQAPAGYANVATHSTLIVLTAQTYNASPVITSSVK-ADRPSL-EKPEPIHLS 2301
QY 1985 -----VPPVSGHATARTPA----- 1999
Db 2302 VSTPVTQGGTVKVLTOGINTPPVLVNVNLVLTESIITNKKLADPVLTKIETKVLQPANL 2361
QY 2000 -KNLAPHASDPDPAPPASADP--HREKTQSKFP--SIQELRLSLGYHGSYSPEGVE 2054
Db 2362 GSTLTTPHH---FPALPSKLPTEVNVHVPSPGSI PADRTVSHLAAAKLDAHSRPSGPGPS 2417
QY 2055 PVSFVSPSLTHDKGL-----PKH---LEELDKSHL- 2082
Db 2418 SFPRASHSPSTASTALSTNATVMAAGIPVPOFISSIHPEQSVIMPPHSITQTVLSHLS 2477
QY 2083 EGELRPKQPG---PVKLGCEAAHLPHLRPLPESQSPSSPLLQTPGVKGHORVVTLAQH 2138
Db 2478 QGEVRNMTPLPSITVSIIRPEALHSR-APL---QP-----QIEVRA-- 2516
QY 2139 ISEVITQDVTNRHPQOLSAFLPAPLYSFFGASCFLVLDLRPPSD---LVLPPPDHGAPAR 2195
Db 2517 -----PQASTPQAP-----AGVPALASQHPPEEVHVLHPVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGGEGDIEPVSPPEGM 2231
Db 2558 SEVLVMQSEYRLHPYTPVRDVRIMVHPHTVAVSEQP-----AADGVVKKVPVASKA 2608
QY 2232 TEPGHSRSAYVLLYRDGEQTEPSRMGSKSPGNTSOP 2268
Db 2609 PQ-----QPGKEAAKTFDAAKAP 2626
```

```
RESULT 29
AAM66665
ID AAM66665 standard; protein; 2665 AA.
XX
AC AAM66665;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26971.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US0000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
(PMOLB-) MOLECULAR DYNAMICS INC.
XX
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PI WPI; 2001-488900/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 26971; 659pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 2665 AA;
```

```
Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

QY 71 EFQPGNRSQELHLRPE-----SHSYLPELKGK-----EMEFIESKR---PRIELLP 114
Db 525 ERKSGQEKSHSVNTEBKIGIDIDHTQSYRKQWQEQSRKQKQWEMEIASKBKFSGPKDV-- 582
QY 115 DPLLRSPLLATQCPAGSEDLTKDRSLTCKLEPVPSPSPHTDPELELVPRLSKELIQ 174
Db 583 DEYERSLVHEVGKP--PDVTD-----SPPSKK-----K 611
QY 175 NMDRVVDREI--TWVEQQISKLKKKQOOLEBEAAKPP-----EPEKPVSPPIETES--- 221
Db 612 RMDHVDFDICTKERNYRSRQISEDSERTGGSPSVRHGSHFDEDEPIGSPRLLSVKGSP 671
QY 222 -----KHSRLVQIIYDENRK--AAAHRILEGGLGPQVLEPLYNQPSDTRQ 265
Db 672 KYDEKVLVPYSNITVRESLSKFNPDSSRRQEQWADMAKIKLSVLNSELNRM-----DSOM 727
QY 266 YHENIKINQAMRKLLIYFKRRNHARKQWKQFCQRYDQJMEALEKKVERIENNPR--- 322
Db 728 KODAGRFVSPFNII-----KRDSLKRKSRVDL--EPGEVPSDSDEGSHKSHSPASAL 781
```


Db 1498 GV-VAVSPKESPOKEDGLSSOLKSDPVDPOKEPEKEDVSASGSPPEATQLAKOMLEBQ 1556
QY 1173 -----SPRGAGPESLGVPTAQASVLRGTALGSV----- 1203
Db 1557 AVEHIAKLAASAAKADAPGLA--PEDRDKPAHQASETELAAGIISIINDISGEPE 1614
QY 1204 -----PGGSITKIGIPSTRV-----PSDSAITYRGSITHGPADVLYKGTI--TRIIGED 1250
Db 1615 NFPAPPYPGESQTDLPAGAAQAPSE-----EGMETDAVSGILETEAATES 1664
QY 1251 S-----PSRLDGRDESLPKGVHIVYEGKKGHVLSEGGMSVTCQSKEDGRSS 1298
Db 1665 SRPPVNAAPPDSAGPTDTKEARGNSSETSHVPEAKGSK-----EVEVTLVRKDKGRQ-- 1716
QY 1299 GPPHETAAKRTYDMWEGRVGRASISIEGLMGRALPERRHSPHLKQEHIRG---SI 1355
Db 1717 -----KTRRRKRNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
QY 1356 TOGIPRSYVEAOBYLRRBAKLLKREGTPPPPPPPSRDLTEAYKTOALGPLKPKAHEGLV 1415
Db 1753 NEGTTVQHPEAQ-----BEKQSEKHPSTPPQSCSTSDLSKIPSTE----- 1792
QY 1416 ATVKEAGRSIHIEPRELKHRT-----PELPLAPRLKESITQGTPLKYDTGASTTGSKK 1470
Db 1793 -----NSSQEISVEERTPTKASVPPDLPPPPQP-----ADVDEPQA-----R 1830
QY 1471 HDVRSLIGSPGRFTF--PVHPLDMDADARALERACEESLKRPGCTASSSGSISARGAPV 1528
Db 1831 FRVHSIIESDPVTPPSDPSIPIPLSV-----TAAKLSPPVAGS-- 1870
QY 1529 IVPCLKPQOSP-----LTVEDHCAPFAGHLPRGSPVTREPTPRLQEGSLSS---SK 1578
Db 1871 -----GIPQSPPTKTEMITRQ-----EPRAQSTPSALPDPTKASDVDTIS 1913
QY 1579 ASQDKLSTPREIAKS-----PHSTVPEHPHPITSPYEHLLRGVSGVDLYR 1625
Db 1914 SSTLRKILMDPKVVSATSVTSTVTTAIAEPVSAAPCLHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDTSIPRGIPLOAAAAYLPHLANPTYPHLYPPLRYGYPDTAALENQIT 1685
Db 1968 KTAPPVTNSETQASVLAADKEKAVPIAKIT-----SVISRMFVSIDLNSQKI 2020
QY 1686 INDYITSQMHNTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLVP 1745
Db 2021 -----TLAKPAQTLTGL-----VSALTGLVNVSLVP--VNALKG 2053
QY 1746 PTFGTATAMDRLAYLPTAPQPFSSRHSSPLSPGGPHTLTKPTTTSSSERERDRDRD 1805
Db 2054 PVKGSVTLTKSLVS-----TPAGFVNVLKGVP----- 2080
QY 1806 RDREREKSLTSTTTVEHAPI-----WRPCTEQSSGS-----SGSS 1841
Db 2081 -----NVLTVGVNVLTTPVNAVTVNAAGPTVNAASAVNATASAVTVTAGAVTAAS 2133
QY 1842 GGGGGS-----SSRPASHSHAHQSPISPRQDALQORPSPVLHNTGMKGIT 1888
Db 2134 GGVATATGVTWAGAVIAPSTCKQKQASANENSRHFGSPMPVIDRPA---DAG-SGAGL 2189
QY 1889 AVEPSKPTVLRSTS---TSSPVRPAA--TFPPATHCPLGGTLDGVYPTLMEPVLPLPKA 1942
Db 2190 RVNTSGVLLSYSGQKTEGPORISAKISQIPPAS-----AMDIEFQQSVSQVKPDS 2243
QY 1943 PRVAPR--ERRADTGHAFIA-----KPPASGLEPASPSPKSEPRPL--- 1984
Db 2244 VTASQPPSGKQAPAGYANVATHSTLVLTAQTVNASPVISVK-ADRPDL-EXPEPHLS 2301
QY 1985 -----VPPVSGHATARTPA----- 1999
Db 2302 VSTPVTQGGTVKVLTOGINTPPVLVHNLVLFPSIVTTNKKLADPVTLKLTETKVLQPANL 2361
QY 2000 -KNLAPHASPPPPAPPASADP--HREKTOQSKPF--STQLELBSLGVHGSYSPEGVE 2054

2362 GSTLTPHH-----PPALPSKLPTEVNVHPSGSPISPADRTVSHLAAAKLDAHSRPSGPGFS 2417
QY 2055 PVSVPSSPSLTHDKGI-----PKH--LEELDKSHL- 2082
Db 2418 SPTRASHPSSSTASTALSTNATVNLAAAGLPVPOFISIIHPQSVIMPPHSITQTVSLSHLS 2477
QY 2083 EGELRPKOPG-----PVKLGEAAHLPLRLPSPESQFSSPLLQTAGPVKGHQRVVTLAQH 2138
Db 2478 QGEVRMNTPTLSITYSIRPEALHSPR-APL---QP-----QQIEVRA-- 2516
QY 2139 ISEVITQDTRHHPOOLGAPLPAPLYSPFGASCPLDLRRPSPD---LYLPPPDHGAPAR 2195
Db 2517 -----PORASTPQAP-----AGVPALASQHPPEEVEVHVLVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGGEGDIEPVSPEGM 2231
Db 2558 SEVLVWQSEYRLUHPYTVPRDVRIMVHPHTAVSEQPR-----AADGVVKVPPASKA 2608
QY 2232 TEPFGHSRAVYLLYRDGEQTEPFSRMGSKSPGNTSQP 2268
Db 2609 PQ-----QPGKEAAKTPDAKAAP 2626

RESULT 31
ABG48336
ID ABG48336 standard; peptide; 2665 AA.
XX AC ABG48336;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID NO 26984.
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX PN WO200157273-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488898/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX PS Claim 27; SEQ ID NO 26984; 658pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. ABG47348-ABG59930 represent human

CC	liver single exon encoded peptides of the invention. Note: The sequence information for this patent does not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	CC
XX	Sequence 2665 AA;	
SQL	Query Match Best Local Similarity 19.3%; Score 518; DB 4; Length 2665; Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;	
QY	71 EFQGNERSQELHLRPE-----SHSVLPDELGKS-----EMEFISKR-----PRLELLP 114	
DB	525 ERKSGQKSHSVNTEEKIGIDIDHTQSYRQMEOSSRRKQKMEMIAKSEKPGSPKDV-- 582	
QY	115 DPLLRPSPLLATGAPGAGEDLTDRSLTGLKLEPVSPSPPHPTDPELELVPPRLSKEELIQ 174	
DB	583 DEYERRSLVHEVGKP--PDQVTD-----SPPSKK-----K 611	
QY	175 NMDVRDREI--TMVEQQISLKKKKQOQLEEEAAKPP-----EPEKVPSPPIBS----- 221	
DB	612 RMDHVDVDICTKRENRNYSRQISSEDSERTGSPSVRHGSHFEDEDPTGPRLLSVKQSP 671	
QY	222 -----KHSLSVQIIVDENRKK--AAAHRIILEGLCPQVELPLYNQPSDTRQ 265	
DB	672 KVDEKVLPSNITVREESLFPNPDYSRREQOMADMAKIKLSVLNSEDENLRW-----DSQM 727	
QY	266 YHENIKNOAMRKKILVFKRRNHARQWKQFCQYDQOLMEALEKKVERIENNPR-- 322	
DB	728 QODAGRFDVFPNSII-----KXDSLRRKRSVRDL--EPGEVPSDDEGEHKSHSPRASAL 781	
QY	323 -----RAKESKVREYVEKQFPIRKQRELQERMQSRVQGRGSL-----SMSAAR 367	
DB	782 YESSRLSFLLRDREDKLREDERLSSSLERNKFYSFALDKTITPTKALLERAKSLSSSR 841	
QY	368 SEHEVSIIDGLSE-----QENLEYQMQLAVIPMLYDADQORTKFINMGCLMADPMKY 423	
DB	842 EEN--WSFLDWDSDRFANFRNKKKEDVDSAPRPPIPSWYMKKKKIR--TDSEGMDDDKKEDH 898	
QY	424 KDRQVMNMWSEQKETREKFMQHPKNGFLTLASLERKTVAECVLYVLTCKENYKSLV 483	
DB	899 KESE-----QERQELFASRPL--HSSIPEQDSKELQ-----HLERKEDSDSFI 940	
QY	484 RRSYRRGKSGQQQQQQQQQQQQQQQQQQ-----MPRSSQEKDEKEKEKEABKEE 533	
DB	941 GRIV--GK--QTSEGANSTDSIQEPVVLFHSRFMELTRMQQKEKEKQKPEVEKQE 994	
QY	534 EKPEVNDKEDLLKEKTDSDTSGEUNDKEAVASGRKTANSQGRRKGRITRSMANBANSE 593	
DB	995 --DTEN-----HPKTPESAPENK-----SELKTP 1017	
QY	594 EAITPQSQASLASMELNESSRWTEEMETAKKGLLEHGRNWSAIARVMGSKTVSOCKNFY 653	
DB	1018 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1046	
QY	654 FNYKKRQNLDEILOQHLKMEKERNARRKKKAPAAASEEAFPPVVEDEMEASGVSGN 713	
DB	1047 -----EAPLVTEEKVTEPATVSEAKPASEPAPA 1075	
QY	714 EEMVVEBAALHAGNEVPRGECGGPATVNNSSDTEISIPSHITEAAKDTGONGKPKPATL 773	
DB	1076 PVEOLEQV-----DLPPGA-----DPDKEAAMP--AGVEEGSSGGQPP--YL 1114	
QY	774 GADGPPPGPTTPPRTRSRAP--IBPTPASEATGAPTPPAPPSFSPAPPPVVPVPEKEEET 831	
DB	1115 DAK-----PPTPGASFQAESNVDPEDS-----TQPLSKPAQKSEANEPEAKKPDAT 1163	
QY	832 AAAPP-----VEEGEEQKPAABEILAVD--TGKAEPEPVKSECTEEAEQPAKGDAEAA 883	
DB	1164 ADAEPDANQAEAAAPESQPPASEDLVDVPPVAAKDKPKNSKRSKTVPQAAAVSIVEKPV 1223	
QY	884 EATAEGALKAEKKEGGGGRATTAKSGSAPQSDS--SATCSADEVDEAFGGDKNRLLSRPP 942	

654 FNYKKQNLDEILQOHLKMEKERNARRKKKAPAAASEEAAAPPPVVEDEMEASGVGN 713
1047 -----EAPLVEETKVPATVSEAKPAPAA 1075
714 EEMVVEEABALHAGNEVREGCSGPATVNNSSDTSIPSPHTEAAKQDQNGPKPATL 773
1076 PVQLSQV -----DLPDGA-----DPDKEAAMP-----AGVEGSSGDQPP-YL 1114
774 GADGPPPPPTPPRRTSRAP--LEPTPASEATGAPTPPAPPSPSPAPPVVPVPEKEEET 831
1115 DAK-----PPTPGASFQAESNVDPEDS-----TOPLSKPAKSEANEPAKPKDAT 1163
832 AAAPP-----VEGEBOQPPAAELAVD---TKABEPVKSECTEABAECPAKGDAEAA 883
1164 ADAEPDANOKAEAPESQPPASEDLEVPVAAAKKPNKSKRSTPVAQAAVSIIVEKPV 1223
884 EATAEGALKAKEGEGGRATTAKSCAPQSDS-SATCSADEVDEAEGDKNRLSPRP 942
1224 TRKSERIDREKLKRSNPRGEAOKLLELMEAEKIRTAASKNSAADLE-----HPEP 1275
943 SL-LTPT-----GDPRANASPOK-PLD-----LKOLKQRAAAALPIPIQVTK 980
1276 SLPLSRTRRNVSVATMGD-HENRSPVKEPVEQPVTRTKRLERLEQAAAVP--TTPR 1332
981 VHEPPR-----EDAAPTKPAPPAPP-----PONL-----QPESDAP 1012
1333 RGRPPKTRRRADDEEENEKAPETLKPPEGWSPRSQKTAAGGGPOGKKGNPKVDAT 1392
1013 QOPGSSPRG-----KSRSPAPPADKFAFAEAQKLPDPCWTSGLPPFPVPREVTKASP 1067
1393 RPEATTEVGPQGVKESSEMEPKAAEEAGSEQQRDRK--AGTDKNPPTAPVVEVVEKPP 1450
1068 HAPDPSAFSAPPGHPLPLGLDHTARVPLPRPTTISNPPLISSAKHPSVLERQIGAISQ 1127
1451 -APEKNSK-----KGRSRSNRLAVDKSASLN-----VDAVSPRGAQAQERES 1497
1128 GMSVQLHVPVSEHAKAPGVVTWGLPLPMDP-----KULAPFSG-----VKQEQOL-- 1172
1498 GV-VAVSPKSESPOKEDGLSSOLKSDPVDPDKEPEKEDVSASGSPSEATOLAKQMELEQ 1556
1173 -----SPGQAGPPESLGVPTAQAEASVLRTGALGSV----- 1203
1557 AVEHIAKLAEASAAKADAPEGLA--PEDRDKPAHQASETELAAGISINDISGEPE 1614
1204 -----PGSITKIGIPSTRV-----PSDSAITYRGSITHGTADVLVYKGTI--TRIIGED 1250
1615 NFPAPPYPYGESQTDLOPPAGAALQOSE-----EGMETDCAVSGILETEATES 1664
1251 S-----PSRLDRGREDSLPKGHVIVEGKKGHVLSEGGMSVTCQSKEDGRSS 1298
1665 SRPPVNPAPPSAGFTDTKEARGNSSETSHSVPEAKGSK-----EVEVTLVRKDKGRQ-- 1716
1299 GPPHETAAKRTYDMMEGRVGRAISSASIEGLMGRALPPEHSPHLEKQHIRG---SI 1355
1717 -----KTRSRKRNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
1356 TOGIPRSYVBAODYLRRRAKLKREGTTPPPPPSRDLTEAVYKTOALGPLKPKAHEGLV 1415
1753 NEGTTVQHEAPO-----BEQSEKPHSTPPQSCSTDLKIPSTE----- 1792
1416 ATVKEAGRSIHEIPRELHRT-----PELPLAPRLKESGITQGTPLKYDTGASTTGSKK 1470
1793 -----NSSQEISVEERTPTKASVFDLPPPPPOP-----APVDEEPOA-----R 1830
1471 HDVRSLIGSPGRTPF--PVHPLDVMADARALERACEESLKRPGTASSGSGSIARCAPV 1528
1831 FRVHSIIESDPVTPPSDPSIPIPLPSV-----FAAKLSPPVAVSG-- 1870
1529 IVPCLKPQOSP-----LTYEDHGAPFAGHLPRGSPVTMRPTPRLOEGSLSS---SK 1578
1871 -----GIPQSPPTKVTWITROE-----EPRAQSTPSPALPDDTKASDVNTS 1913
1579 ASQDRKLTSTPREIAKS-----PHSTVPEHHPHIPISPYEHLRIGVSGVDLYR 1625

1914 SSTLRKILMDPKYVSATSVTSTVTTAIAEPVSAAPCLHEAPPPVD-----SKPLEE 1967
1626 SHIPLAFDPTSIPRGIPLDAAAAAYLPHRLAPNPTYPHLYPPVLYIRGYPDPTALENRQTI 1685
1968 KTAAPPVNNSEIQASEVLVAADKEKVPVIAPIKIT-----SVISRMVPSIDLENSOKI 2020
1686 INDYITSQMHNTATAMAOADMLRGLSPRESSLALNAAAGPRGIDLSQVPHLPVLP 1745
2021 -----TLAKPAQTLTGL-----VSALTGLVNVSLVP-VNALKG 2053
1746 PTFGPTATMDRLAYLPTAPQPFSSRHSSPSLSPGQTHLTKPTTTSSSERERDRDRD 1805
2054 PVKGSVTTLKSLVS-----TPAGFVNVLKGVP----- 2080
1806 RDREREKSLTSTTTTVEHAPI-----WRQTEQSSGS-----SGSS 1841
2081 -----NVLTPGVNVLTTTPVNAATVGTVNAAPGTVNAASAVNATASAVTVTAGAVTAA 2133
1842 GGGGGS-----SSRPASHSHAHQSPISPRQDALQORPPSVLHNTGMKGIIT 1888
2134 GGVATTGTVTWAGAVIAPSTCKQKQASANENSRPHGSMPIVDDRPA---DAG-SGAGL 2189
1889 AVEPSKPTVLRSTS---TSSPVRPAA---TFPPATHCPGLGTLGVYPTLMEVLLPKEA 1942
2190 RVNTSEGVLLSYSGOKTEGPORISAKISQIIPAS-----AMDIEFQOSVSKSQVKPDS 2243
1943 PRVARP--ERRPADTCHAEFLA-----KPPARSGLPEPASSPSKGSPPRL--- 1984
2244 VTASOPPSKGPQAPAGYANVATHSTLVLTAQYNASPVISSVK-ADRPUL-EKPEPIHLS 2301
1985 -----VPPVSGHATIARTPA----- 1999
2302 VSTPVTQGTVKVLTOGINTPPVLVHNLVLTPSIVTTNKKLADPVTLKTIETKVLPANL 2361
2000 -KNLAPHASPPPPAPPASDP--HREKTQSKPF--STQELERLSLGVHSGSSYSEGEV 2054
2362 GSTLTTPHH---PPALPSKLPTENVHVPSPGSIPTADRTVSHLAAAKLDAHSPPSPGPGS 2417
2055 PVSVPSSPSLTHDKGI-----PKH--LEELDKSHL- 2082
2418 SPFRASHPSSTASTALSTNATVNLAAAGIPVPOFISIIHPEQSVIMPPHSITQTVSLSHLS 2477
2083 EGELRPKQPG-----PVKLGEAAHPLHPLRPESQSSPPLLQATAPGVKHQVVTVAQH 2138
2478 QGEVRMNTTLPSTIYSIRPEALHSPR-APL---QP-----QQIEVRA-- 2516
2139 ISEVITQDVTTRHHPOOLSAPLPAPLYSPFGASCPLDLRRPPSD---LVLPPPDHGAPAR 2195
2517 -----PORASTPPQAP-----AGVPALASQHPPEEVHVLHPLVARATAPVQ 2557
2196 GS-----PHSEGGKRSPEPNKTSVLGGEDGIEPVSPPEGM 2231
2558 SEVLVWQSYRULHYTYTPRDVRIMVHPHVTAVSEQPR-----AADGVVKVPPASKA 2608
2232 TEPGHRSRVAVYLLYRDGEGTEPSRMGSKSPGNTSQP 2268
2609 PQ-----QPGKEAAKTPDAKAAP 2626

RESULT 33

ABG36319

ID ABG36319 standard; peptide; 2665 AA.

XX ABG36319;

XX ABG36319;

XX 19-AUG-2002 (first entry)

DT Human peptide encoded by genome-derived single exon probe SEQ ID 25984.

DE Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagenier syndrome;
pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
primary ciliary dyskinesia; pulmonary hypertension;
hyaline membrane disease.

Homo sapiens.
WO200186003-A2.
15-NOV-2001.
30-JAN-2001; 2001WO-US000665.
04-FEB-2000; 2000US-0180312P.
26-MAY-2000; 2000US-0207456P.
30-JUN-2000; 2000US-00608408.
03-AUG-2000; 2000US-00632366.
21-SEP-2000; 2000US-0234687P.
27-SEP-2000; 2000US-0236359P.
04-OCT-2000; 2000GB-00024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2002-114183/15.
Spatially-addressable set of single exon nucleic acid probes, used to
measure gene expression in human lung samples.
Claim 27; SEQ ID NO 25984; 634pp; English.

The invention relates to a spatially-addressable set of single exon
nucleic acid probes for measuring gene expression in a sample derived
from human lung comprising single exon nucleic acid probes having one of
12614 nucleic acid sequences mentioned in the specification, or their
complements or the 12387 open reading frames derived from the 12614
probes. Also included are a microarray comprising the novel set of probes
; the novel set of probes which hybridise at high stringency to a nucleic
acid expressed in the human lung; measuring gene expression in a sample
derived from human lung, comprising (a) contacting the array with a
collection of detectably labeled nucleic acids derived from human lung
mRNA, and (b) measuring the label detectably bound to each probe of the
array; identifying exons in a eukaryotic genome, comprising (a)
algorithmically predicting at least one exon from genomic sequences of
the eukaryote; and (b) detecting specific hybridisation of detectably
labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
having a fragment identical to the predicted exon, the probe is included
in the above mentioned microarray; assigning exons to a single gene,
comprising (a) identifying exons from genomic sequence by the method
above and (b) measuring the expression of each of the exons in several
tissues and/or cell types using hybridisation to a single exon
microarrays having a probe with the exon, where a common pattern of
expression of the exons in the tissues and/or cell types indicates that
the exons should be assigned to a single gene; a peptide comprising one
of 12011 sequences, mentioned in the specification, or encoded by the
probes/open reading frames (ORF). The probes are used for gene expression
analysis, and for identifying exons in a gene, particularly using human
lung derived mRNA and for the study of lung diseases such as asthma, lung
cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
Karsagenier syndrome, fibrocystic pulmonary dysplasia, primary ciliary
dyskinesia, pulmonary hypertension and hyaline membrane disease. The
present sequence is a peptide/protein encoded by a single exon probe of
the invention. Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ	Sequence 2665 AA;	
Query Match	3.9%; Score 518; DB 5; Length 2665;	
Best Local Similarity	19.3%; Pred. No. 9.2e-20;	
Matches 505;	Conservative 315; Mismatches 863; Indels 934; Gaps 120;	
QY	71 EFQGNERSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR-----PRLELPP 114	
DB	525 ERKSGQEKSHSVNTEEKIGIDIDHTQSYRKQWQSRKQKQWEMETAKSEKFGSPKKDV-- 582	
QY	115 DPLLRPSLLATGQAGSEDLTKDRSLTGKLEPVPSPSPPHDTDPLELVPPLSLKEELIQ 174	
DB	583 DEYERRSLVHEVGKP--PDQVTD-----SPPSKK-----K 611	
QY	175 NMDRVREI--TWVEQOISKLKKQOOLBEEAKPP-----EPEKPVSPPIES----- 221	
DB	612 RMDHVDFTCTKRNRYSSROISDSERTGSGPSVRHGSHFDEDDPGISPRLLSVKQSP 671	
QY	222 -----KHSILVQIYDENRKK--AEAHRILEGLGPQVELPLYNQPSDTRQ 265	
DB	672 KYDEKVLVPSNITVREESLKFNPYDSRRREQWADMAKIKLSVLNSEDELNRW----DSQM 727	
QY	266 YHENIKINQAMRKKLILYFKRNHARKQWKQFCORYQOLMEALKKQVERIENRPR--- 322	
DB	728 KQDAGRFDVSPFNII-----KQDSLTKRSVRDL--EPGEVPSDSDEDEGHHKSHSPRASAL 781	
QY	323 -----RAKESKVRVEYKQFPEIRKQRELQERMQSRVQGRGSL-----SMSAAR 367	
DB	782 YESSRLSFLLRDRDKLRDERLSSLSLERNKFYSFALDKITTPDTKALLERAKSLSSR 841	
QY	368 SEHVESEIIDLSE-----QENLEKQMRQLAVIPPMYLDADQOQRIKFINMNGLMADPMKY 423	
DB	842 EEN--WSFLWDMSRFANFRNNKKEKVDSPAPRIPSWYMKKKIR--TQSEGMDDKKEDH 898	
QY	424 KDRQVNNWSEQEKETFRFKPMOHKPNGLIASFLERKTVAECVLYLYLTKNVYKSLV 483	
DB	899 KEEE-----QERQELFASRFL-HSSIIEQDSKRLQ-----HLERKEEDSDFTS 940	
QY	484 RSYRRRGKSGOQQOQQOQQOQQOQQOQ-----MPRSSQEEKQEKQEKQEKQEK 533	
DB	941 GRIY---GK---QTSSEGANSTTDSIQEPVLFHSPFMELTMOQKEKQEKQEKQEK 994	
QY	534 EKPEVENDKDLLEKTDOTDGTGEDNDEKAVASKGRKTANSQGRKRGKITRSMANESE 593	
DB	995 ---DTEN-----HPKTPESAPEND-----SELKTP 1017	
QY	594 EAITPQQAELASLMELNERRWTEBEMETAKGLLEHGNWSAIARMVGSKTIVSOCKNFP 653	
DB	1018 PSVGPPSV-----TVVTLSSAP-----SALEKTTGDKTV----- 1046	
QY	654 FNYKKRONLDEILOQHLKMEKERNARRKKKAPAAASEEAAFPVVEDEMEASGVSGN 713	
DB	1047 -----EAPLVEKTVETPATVSEAKPASEPAPA 1075	
QY	714 EEMVVEEABALHASNEVPRGCSGPATVNNSSDTESIPSPHTEAAKDTGQNGPKPATL 773	
DB	1076 PVEQLEQV-----DLPPGA-----DPDKAAMMP-----AGVBEGSGDQPP-YL 1114	
QY	774 GADGPPPGPTTPPRTSRAP--IEPTASEATGATPPAPPSAPPVVPVPEKEEET 831	
DB	1115 DAK-----PPTPGASFQAEENVDPEDPS-----TQLSPQAQSEANEPKSKPDAT 1163	
QY	832 AAAPP-----VEEGEQKPPAAEELAVD---TGKAAEEPVKSECTEABEGPAKQDAEAA 883	
DB	1164 ADAEPDANKQAEAPESPQPADEDLEVPVPAADKKFNKSKRKTTPVQAAAVSIVEKPV 1223	
QY	884 EATAGALKAEKKEGGSGRATTAKSSGAPQSDS--SATCSADEVDEAEGDKNRLLSRP 942	
DB	1224 TRKSRIDREKRLKSNSPRGEAQKLELMGEAEKTRTASKNNSAADLE-----HPEP 1275	
QY	943 SL-LPPT-----GPRANASPK-PLD-----LKOLKRAAAIPIQVTK 980	
DB	1276 SLPLSRTRRRNRVSVYATMGD--HNRSPVKPEVQPRVTRKLELRELQEAANAAP--TTPR 1332	

Qy	2316	MPALTG--TCLMTYRSGAQVEHASTNMGLEAIRKALMGKYDOWEESPPLSANAFNPLN	2372
Dd	1487	RPOIYGAWGSTQQYRPQPSSPAPQN-----WGAPP---RGAAEPGG	1527
Qy	2373	ASASLPAAMPITTAADGRSDTLTSPGGGKK-----AKVSRGPSRKKA	2414
Dd	1528	APHGPPIQP--AGVAQWDQHRYFPQGPPPPPQQQQPQQQQQQPYQVAGPFGQQPP	1585
Qy	2415	KSPA-----PGLASGD-----RPPSVSVSHSEGDCNRRTPLNTRVWEDRPSAGS	2459
Dd	1586	QAPPQWAQMNPQTQAQGIAPPGGSLRPPS-----GFCQQNRMPGMAQ--QQQSQQQGG	1638
Qy	2460	TPFF-----YNPLIMRLQAGVM-----ASP PPP-----GLPAGSGPLAGPHHAWE	2500
Dd	1639	VPOPPQQAASHGGVPSPCLPVQPGGVKWKPPYAMPPPPSQGVGOOVQGGP---PCGMMSQ	1695
Qy	2501	BPKPL	2505
Dd	1696	KPPM	1700

RESULT 35

ABG70019

ID ABG70019 standard; protein; 2703 AA.

AC ABG70019:

05-NOV-2002 (first entry)

DE Larval viability associated protein #18.

Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;
KW
rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar be
KW
oilseed rape; soybean; vegetable crop; fruit.
KW

OS *Drosophila melanogaster*.AA
PN
WO200257455-A2.

25-JUL-2002.

18-JAN-2002: 2002WO-US001568-XX
PF

18-JAN-2001: 2001US-0262351P.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG-

PI Stam L. Bachmann J. Broadus J. Kamdar KP:

WPI: 2002-590745/63.

DR N-PSDB; ABS51395.

Identifying inhibitors of activity of proteins essential for *Drosophila* larval viability comprises expressing in a host a protein essential for larval activity and identifying compounds that inhibit or interact with the protein.

PS Claim 1; Page 108-117; 169pp; English.

CC The invention describes a method of identifying compounds that inhibit
CC the activity of, or that interact with a protein essential for Drosophila
CC larval viability comprising expressing in a recombinant host a DNA
CC molecule to produce a protein essential for larval viability. The method
CC is useful for identifying compounds with insecticidal activity. Compounds
CC identified are useful as insecticides in crops such as maize, wheat,
CC oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar
CC beet, oilseed rape, soybeans, vegetable crops and fruits. This is the
CC amino acid sequence of a fruit fly larval viability associated protein

Sequence 2703 AA;

Query Match

Query Match 3.9%; Score 512.5; DB 5; Length 2703;

Best Local Similarity 20.1%; Pred. No. 1.9e-19;

Matches	435;	Conservative	187;	Mismatches	746;	Indels	797;	Gaps	103;		
Qy	670	KLXMEKERNARRKKRKKAPAA-----ASEEAAFPVVVEDEMEASGVSGNEEMVBEAEA	723	Db	4	KIKSPOTQQOQGGAAPAATPFSAGAAPGCAATPPT-----	39	Qy	724	LHAGSNEVPBGCSGPATPVNNSDTSIPS-----PHTEAAKDGTQNGPKPPATLG	774
Db	:	: : :	:	Db	:	: : :	:	Db	40	-----SGPPTNNNSNGSDPSIQOOQNVAHPHYGA-----PPPFGS--	77
Qy	775	ADGPPPGP-----PTPPRRTR-----APIBTPASEATGAP-----	806	Db	78	GPCGGPGDPAAVWHYHLHQOQQHQPpppphmqqqhgggpaPPPGGAPGHAPGVKKEY	137	Qy	807	--TTPPAP-----PSPSAPPVVPVKEEKEETAAAPVVE-----GE	841
Db	:	: : :	:	Db	138	THLPpphpPAYGRYHADPNWDVRYCQLPGCKKP-----QQQCPHPQOQPPQPGPGGS	193	Qy	842	EQRPPAAELAVDTGKAEEVPKSECTEEAEEGP-----AKGDAEAETAAGALKAEK	895
Db	:	: : :	:	Db	194	PNRPPOORYTFGOPPQCGTPTLNLSLOSSNPFPPOHRVANTYDPOAAAASAAAAAQO	253	Qy	896	KEGSGSPAITAKSSGAPODSATSATCADEVDEAGDKN-----RLLSpr-----	941
Db	:	: : :	:	Db	254	QQAG-GPPPHGCHGPPPPQHOPS-----PYGGQCGGWAPPRPYSPLGLGPSQOYR	301	Qy	942	--FSLTLPTGD---PRAN-----ASPOKLDLKLQKRAAAI-----PPIQVT	979
Db	:	: : :	:	Db	302	TPPTNTSRGOSYPYPAHGQNSGYSPSSFOOOQQOQQOQQOAGQGGVPGGPPGPGTG	361	Qy	980	KVHEPPREDAAATKPAAPPAP---PPQNLL-----QPEDAPQFGSSPRG	1021
Db	:	: : :	:	Db	362	Q--QPQONTPTPTSQYSPYQRYPTPLGAGGSNHRTAYSTHOYPEPNRMPCGSSPSP	419	Qy	1022	KRSRAPPADKEAFAAEAQKLCGDPPCWTSGLFPVPPPREVIKASP-HAPDPSAF-SYAP	1079
Db	:	: : :	:	Db	420	GSGHLPUPPASPH-HVPPLOQQPPPPHVHSAGGP---PP-----SSSCHAPASPSQPQSQA	471	Qy	1080	PCHPLPL-----GLHD---TARPVLRPPPTISNPPPLISSAKHPSVLERQIGAI	1125
Db	:	: : :	:	Db	472	PSPHQELIGQNSNDSSGGNAHGMGSGPGCTFNPOQVMRPTPSPTCSS-----GSR	522	Qy	1126	SQCMSVOLHVYPYSEHAKE-----APV--GPVTWGLPLPMDPKKLAPFSVKQEQSLSPRGQ	1177
Db	:	: : :	:	Db	523	SMSPAVAQNHPISRPNASNQSSSGPMQOPPVGAGGPPMPHPHGM-P-GGPPQQQSQOQQQ	581	Qy	1178	AGPPESLCVPTAQEASVLRGTALGSVPGSGITKGIESTRVPSDSAITYRGSITHGTPTADV	1237
Db	:	: : :	:	Db	582	ASNSASASANSPOOT-----PPAPPPNQGMNMNMTATPPPPP---QGAAGGYMPP	639	Qy	1238	LYKGTITRIIGEDPSRLDRGEDSLPKGHVIVEGKGHVLSYEGGMSVTCQSKBEDRSS	1297
Db	:	: : :	:	Db	630	HMGGYK--MGFGQGFAGCYPPQPOQ-----QYPPGNYPVRPQYPPGAYA	674	Qy	1298	SGPPHETAAPKRTYDMMEGRVGRAISASIEGLMWRAIPPE--RHSPHHL---KEQHHR	1352
Db	:	: : :	:	Db	675	TGPP-----PpptsQAGAGGANGSNPSCQAAGGYPGRGMNHTGOYPPYOWVPPSPOOTVP	729	Qy	1353	GSITQIGIPSYBAQEDYLREAKLLKREGTppppppSRDLTAYKTQALGPLLKPAHE	1412
Db	:	: : :	:	Db	730	G----CAPGGAMVGNH-----VOGKTppppPVVG-----GP-----PPPQ	760	Qy	1413	GLVATVKEAGRSIHETPRELHRTPELAPRALPKESITOGTPLYKDYTGASTTGSKKH	1472
Db	:	: : :	:	Db	761	G-----SGSPRPNLNYL-KQHLOHKGGYGSGFTP-----PQG-POGYGNG--PTGM----	801	Qy	1473	VRSLIGSPGRTFFPVH-----PLDVNADARALERACYEESLKSREGTASSSG	1519
Db	:	: : :	:	Db	802	-----HPGPMWGPFHHMGPPHGTNNMGPTSTPQSOQLQ-----GGQPOGQGASG	847	Qy	1520	GSTARAGAPVIVELGKPRQSLTYEDHGAFFAGHLPRGSPVTVRETPTRLQEG--SLSSS	1577
Db	:	: : :	:	Db	848	G-----PSSGGEH-----ISQNGTSSSG--PTGA-AGHMAVTSVVTTTGPDCGTSM	891				

Db 1618 TKALGLE-----BSLVQEGR--AREQBEKWR-----CQDVV-----QEW 1650
QY 610 NESSRWTEBEMETAKKLEHGNWGAIAARMVSGKTVSQCKNFYFNKKRQNLDEILOOH 669
Db 1651 QETSPTRPAGBQK-----ELAPAW-----EDTSPQDNRYWGRGDEVALEQDTYWR 1698
QY 670 KLQWERNARRKKKAPAAASEAAFPVVEDEMEASGVSGNBEEMVEEABALHAS-- 727
Db 1699 ELSCE-----RKWVFPHELDQGGARPHYTEERESTFLDEGDDQEVPLRE--HATRS 1749
QY 728 -----GNEVPRCEGSGPATVNNSSDTE-----SIPSPHTEAAKDTGN 765
Db 1750 PHASDFKQESSFQKLEVERWLAESPVLPEEEDKLTSPFEILSPASPPEWVGQR 1809
QY 766 GKPPATLADGGPPGPPPTPPRTSRAPTEPASEATGAPTP-----PAPPSPSAPP 819
Db 1810 VFSAP---GOESPDPKLMFMKN-----EPTTPSWL--ADIPWVPKDRPLPAPLSA 1860
QY 820 PVVPKEEKEETAAAPVVEGEQKPPAAABELAVDTGKABEVPKSECTEAEAGPAK--G 877
Db 1861 PGPTTAPESHTPA--PFSWGTAEDYSV--AAVQEGAA-----ELEGGPSPLG 1906
QY 878 KQAEAAEAATAGALKAKKEGGSGRA--TTAKSSGAPQSDSDSACSADEVDEAEGDXN 935
Db 1907 KQYKAE-----GEREEGRABADKSSHSKVPKASKSHATTEPQTE-----PEQ 1953
QY 936 RLISPRPS-----LILT-----PTGDRANASPOKPP--LDLKQL-----966
Db 1954 REPTPYDERSFOYADIEQOMLTGLGPACTREPPPLGAGDWPPCLSTYKEAAGRNTSA 2013
QY 967 -KORAAAIPIQVTKVHEPREDAATKP--APPAPPPQNLQSPESDAPQPGSSSPKGR 1024
Db 2014 EKELSPISGP-KSLQSDTPTFSVAALAGTVPVPRPFGSPMEPLTTPAVPPRAPI--LSK 2071
QY 1025 SPAPPADKAFAREA--QKLPDGPWCWTSGLPFPVPREVIKASPHADPDASFAYAPGHP 1083
Db 2072 GSPPLNGILNCSPPRRSPSKESGRSHWDDSTSDSELEKGAHQEKEAQSPSP--HP 2130
QY 1084 LPLGLHDTARVPLPRPPTTINPPPLISSAKHPVLERQIGAISQGMVOLHVPYSEHAKA 1143
Db 2131 IPWG-----SPTL-WPETAHVSPPLDHLGPA-----RPSLDFFASAFGS 2171
QY 1144 PVGFVTMGLPLMDPKKLAP-----FSGVKQEQLSPRQAGPP-----ESLGVPVTAQEA 1192
Db 2172 SLOPAPPQLPSPAEPSPS-APCGSLAFSGDRLALAP---GPPTRTRHDEYLEVTKAPSL 2226
QY 1193 -SVLRGTALGSPGGSITKGIPTSTRVPSDAITVYRGSITHGTADVLVYKGTITRIIGDS 1251
Db 2227 DSSLPOLPSPSPGAPLLNLPR---PASPALSEGSSSEATTP-----VISSVAERFS 2276
QY 1252 PS-----RLDRGREDSLPKGHVIEYEGKGHVLSYEGGMSVTQCKSKEDGRSSGPPH 1302
Db 2277 PSLEAAEQSGELDPGME---PAHSLWD-----LTPLS-----2307
QY 1303 ETAAPKRTYDMMEGRVGRAISSA-SIEGLMGAIPP-----ERHSPHLKQHHI 1351
Db 2308 --PAPPASLDL-----ALAPAPSLPGMDGILPCHLECSAAATEKSPFPQVPSDECA 2358
QY 1352 RGSITOGIPRSYAEQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQALGPLKLKPAH 1411
Db 2359 ANGPTETSP-----NPPGAPAKAENE-----2380
QY 1412 EGLVATVTKAG--RSIHEIPELRLHTPELPLAPRPLKSGSITQGTPLKYDTGASTGS 1468
Db 2381 EAAACPAWENGAWPEGAERSSRPDTLLSPQVCPAGGSGPPSSASP-EVEAGPOGCAT 2439
QY 1469 KKHVRLSLGSPGRTFPFPHPLDVMADARALERACYEESLSKSPGTASSGGSIGARGPV 1528
Db 2440 EPRPHRGEL-SPSFLNPLPP---SIDDRDL---STEEVRLVGR-----GRRRRVGPG 2486
QY 1529 IYPELCKPQSPLYEDHGAFFAGHLPRGSPVTWREPTPLQEGSLSSSKASQDKLTST 1588
Db 2487 TT-----GGPCPVTDETPTTSASDSGSSQSDSD-----V 2515

RESULT 38

AAB50362

ID AAB50362 standard; protein; 3118 AA.

XX AC AAB50362;

XX DT 12-MAR-2001 (first entry)

XX DE Human SRCAP.

XX KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
KW CAMP regulatory element; CREB binding protein; CBP; Atpase;

XX KW transcription activation; DEAD box RNA dependent helicase;

XX KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.

XX OS Homo sapiens.

XX PN WO200073467-A1.

XX PD 07-DEC-2000.

XX PF 25-MAY-2000; 2000WO-US014719.

XX PR 27-MAY-1999; 99US-0136620P.

XX PR 25-MAY-2000; 2000US-00579181.

XX PA (UYSL-) UNIV SAINT LOUIS.

XX PI Chrivia J, Yaciuk P;

XX DR WPI; 2001-061545/07.

XX DR N-PSDB; AAC89859.

XX SNf2 related CAMP regulatory element (CREB) binding protein (CBP)
XX activator protein, capable of co-activating CREB binding protein, useful
XX for modulating transcription and for affecting viral infection.

XX Claim 6; Page 77-86; 103pp; English.

XX The present sequence is an Snf2 related CREB (cAMP regulatory element)
XX binding protein (CBP) activator protein (SRCAP) polypeptide. It has
XX Atpase activity and is capable of activating transcription. SRCAP
XX polypeptides are useful for activating transcription in a cell, for
XX enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
XX activation of transcription in a cell, for treating a patient having a
XX disease involving a function such as insufficient transcription of a
XX gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
XX helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
XX as affected by SRCAP protein. Compounds that modulate SRCAP function, such
XX as antibodies, antisense molecules, polynucleotides or ribozymes, are
XX useful for treating diseases mediated by SRCAP-activated transcription,
XX for example, infection by adenovirus, hepatitis C virus, human
XX immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
XX hepatitis B virus

[illegible]

1676 -TAALENRO-----TIINDYTSQ---OMHNTATAMAQADMRLG- 1712
1933 QTLAVLLRLQKABHRVLI FTOMTMDLVLEQELTVHGHLVLRDLGSTRVEQOALMERF 1992
1713 -----LSPRESSALNAAAGPRGIIDLSQVPHLPVLVPPPTCTPATAMDR----- 1757
1993 NADKRIFCFLSTRSGGVNLTGADTVVFYDSW-----NPTMDAQADCHRG 2043
1758 -----LAYLPTA-----POPFSS 1770
2044 QTRDVHIYRLISERTVEENILKKNQKRMGLDMAIEGNGFTTAYFKQOTIRELFDMPLEE 2103
1771 RHSS-PLSP-----GPHLTKPTTSSSERDR----- 1800
2104 PSSSSVPSAPEEBEETVASKQTHILEQALCRADEEDIRAATAQAKQVAELAEFNENDG 2163
1801 -----DRDRDREREKSIILSTTTVEH----- 1823
2164 FPAGEGEAGRPCAEDDEMSRAQETAAALVEQUTPIERYAMKFLASLEVSREELKQAE 2223
1824 -----APIWR-PGTEQSSSGS--SSGGGGSSSRPASHSHAHQH--SPIS 1864
2224 EQVEAARKDLDQAEKEVFRLPQEEEGPGAGDESSCGTGCTHRRSKKAKAPERGTRVS 2283
1865 PRTQDALQORPSVLHTGMKGIIITAVEPSKPTVLRSTST---SSVPRPAATFPPTHCP 1921
2284 ERLRGAEATQGANHTP-----VISAHQ-----TRSTTTTPRCSPARERVPREPAPRPT 2333
1922 GGTLDGVYPTLMEPVLLPKAEPVARPERPRAUTGHAFKAPPARSGLEPASPSKSEP 1981
2334 PASAPAAIPAL-----VVPVSAVPVPIGAPNPTITILPVHILSPPPSPQIPCSSPA--CTP 2388
1982 RPLVPPVSGHATTIARTPAKNLAPHASP---DPPAPPASAS-----DPHREKTSKP 2030
2389 PPACTPPAHTP---PPAQTCLVTPSPPLLGPPSPVPIASVTNPLGLRPEALCAQAL 2445
2031 FSQLELELSLGVHGS-----YSPEGVSVS-----PV-----SSPSLTHDKG-LPK 2072
2446 ASPSELELASVASSSETSSLSLVPKOLLPVAVEILPVSEKNLSLTPSAPSLTLEAGSIPN 2505
2073 HLELOKSHLEGLRPQCPVKLGGEAHL-----HLRPLPESOPSSPLLQTAGV 2126
2506 GQEQEAPDSAGTTLVLP-----EGEELPLCVSENGLELPPPSAASDEPLEAD 2558
2127 KGHQVVVTLAQHSIVITDYTRHHQQL--SAPLAPLVSFPGASCPVLDLRRPPSDLYL 2185
2559 R-----TSEELTEAKTPTSSPEKQELVTAEVAAPSTSSATSP----- 2598
2186 PPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGEDGIEPVSPPEGMTEPGHRSKSAVYPLL 2245
2599 -----EGPSPARPPR----- 2608
2246 YRDGEQTEPSRMGSKSPGNTSOPPA--FFSKLTESNMAVSKKQKINKLNTNRNEPE 2303
2609 RRTSADVEIRGQGTGRPG---QPGPKVRLKPLGRILVTVVEEKELVORR----- 2655
2304 YNISQPGTEIFNPAITGTGLMTYRSQAOVHAETNMGLEAIIRKALMGKYDOWESSPLL 2363
2656 ---QQRGAASLTVPVSET-----SASPGSPSV--RMSGP-----ESSPI 2692
2364 S-----ANAFNPLNASLPAAMP--ITAADGRSDH 2392
2693 GGPCEAAPSLSLTPPQQPIARRHIELVGTGGSPENGDGALLAITPAVKRRRRGRPK 2752
2393 TLTSPPGGGKAKVGRPSRKAKS-----PARG-----LASDRPPSVSSH 2434
2753 KNRSPADAGRGVDEAPSSSLTKGTNGADVPVGPETIIVADPVLEPOLIPGQPLGPQVH 2812
2435 SEGDCNRRTPLTNRVWE---DRPSSAGSTPPFYNPLMRLOAGWAS-----PPPP 2482
2813 -----RPNPLLSVPEKRRGRPPKARDLPIP-----GTISAGDNGSESTQPP 2856
2483 GLPAGSGLAGPHHAWDEEPKPLL 2507

Db 2857 PHPSPLTPL-----PPLVC 2871

RESULT 39

ABB62899 standard; protein; 3201 AA.

AC ABB62899;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 15489.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL07002.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

PS Disclosure; SEQ ID NO 15489; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ffp.wipo.int/pub/published_pct_sequences

XX Sequence 3201 AA;

Query Match 3.6%; Score 473; DB 4; Length 3201;

Best Local Similarity 18.6%; Pred. NO. 4e-17;

Matches 573; Conservative 378; Mismatches 1132; Indels 994; Gaps 130;

QY 77 ERSQELHLPESHVLPGLKSEMEFIESKRPRELLLPDLRPPSPPLATGPAGSEDLT 136

Db 183 EKGVQVNVNPDSS---LPE-----PCVYVLPPEVW-----SPAHDKLP 217

QY 137 KORSITKLEPVSPSPPHPTDPELELVP--PRLSKEELIQNMDRVREITMVQQISKLK 194

Db 218 KNYL-----RITKDEHYSPLDLHVTNVVALAENTCAYDIDPID-----EAWRLYN 264

QY 195 XKQQQLLEEAAKPEPEKPVSPPPPIESKHSRLVQIIYDENRKKAAHRILEGIG----P 250

Db 265 SDRAOQGAPFINATQERVIE--ELEVRQWEQIQVIL-----KLEEGIGIEFDE 311

QY 251 QVELPLYNOPSTRQVHE-----NIKINQA-----MRKKLLY 283

Db 312 NVICDVCRSP-DSEANEMVFCNDCNICVHOACYGITAIPSGQWLCTRCMSGKPCVLC 370


```
QY 2471 LQAGVMSPP-----PPGLPAGSGPLA 2492
: : : : :
Db 2328 TKSTTTSDPBNICKVKPQQLQTSLSPSA 2355

RESULT 41
ID AAB35408
XN AAB35408 standard; protein; 2819 AA.
AC AAB35408;
XX
DT 23-MAY-2001 (first entry)
DE Human 07CG27 gene protein.
XX
KW Human; 07CG27 gene; chromosome 1; HPC1 region; prostate cancer; oncogene.
XX
OS Homo sapiens.
XX
PN WO200116291-A2.
XX
PD 08-MAR-2001.
XX
PF 25-AUG-2000; 2000WO-US023291.
XX
PR 27-AUG-1999; 99US-0151049P.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI Tavtigian SV, Swedlund B, Simard J, Rommens JM;
XX
DR WPI: 2001-226682/23.
DR N-PSDB; AAF28060.
XX
PT Novel human prostate cancer marker gene termed as 07CG27 gene, useful for
PT screening mutations in the gene in diagnosis of a predisposition to
PT cancer.
XX
PS Claim 1; Page 91-99; 99pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC human 07CG27 oncogene. This gene is found at the HPC1 region of
CC chromosome 1. The sequences can be used in the diagnosis and
CC identification of treatments for prostate cancer. The present sequence is
CC the 07CG27 protein
XX
SQ Sequence 2819 AA;

Query Match 3.6%; Score 469.5; DB 4; Length 2819;
Best Local Similarity 19.1%; Pred. No. 5.3e-17;
Matches 532; Conservative 332; Mismatches 980; Indels 944; Gaps 130;

QY 163 VPRLSKEELIONMDRVDREITWBOQISKLKKQOQLEBAAK--PPEPEK--VSPPP 218
: : : : :
Db 56 MPPANPLSLKAENKGNPNVIVPKDGTWASKQEQHEEKEPTVPVPAQPKGVAAPE 115
: : : : :
QY 219 IESKHSRLVQIYDNRKKAABAHRIEGLGPV-----ELPLYNQPSDTROYH----- 267
: : : : :
Db 116 VAPAPKS-----WASNKQGG-----QDGIQVNSQFQEFPSLQAAGDQEKKEKETND 163
: : : : :
QY 268 ENIKINOAMRKLLILYFKRRNHA-----RKQWKQF-----CORVDOLMEALEKKV 313
: : : : :
Db 164 DNYGPPSLPPNPNVACWDGKGAAGSPSSDQDEKLPQDESTAGTSEQNDILUKVVEKRI 223
: : : : :
QY 314 -----ERIENPRRAKESKVR-----EYEEKQFPEIR-----KQRELQERMQRVGQGS 359
: : : : :
Db 224 ACGPPQAKLNGQQAALASQYRAMPPYMFQYPRMTYPLHLGPMRPPPSLSETKGLRGR 283
: : : : :
QY 360 GLSMSAARSHEVSEIIDGLSEQENLEK-----OMQLAVIPMLVDADQ 404
: : : : :
Db 284 GPPPSWA-SEPERPSILSA-SELKELDKFNDLDAEADGAGQMEVDYTEQLNFSDDDD 341
: : : : :
```


Qy	1282	GMGSMVTCQSKEDGRSS-----SGP-----PHTAAPKRTYDMMEGRVGRALIS-----	1322
Db	1221	-----POYDNKPRAEHIPSGLRQRESESSDFSGVFPVRRRQRGSETDSEIHE	1275
Qy	1324	SASIEGLMGRALIPERH-----SPH-HLKEQHHIRGSITQGI:PRSVV-----BAQED	1369
Db	1276	SASDKOSLSKGKULPKREERPENKKPVKPHSGSPKPDNNHVRID-NRLLEKPYVRDDDKAKPG	1334
Qy	1370	YL-----BREAKLKREGTPPPPPPSRDLT---EAYKTQALGPKLKLPAHEGLVATVK	1419
Db	1335	FLPKGEFTRGRGGTFRGGRDGGRRSPRSTLRRPAYRDNQNNPQO-----	1381
Qy	1420	EAGRSIHEIPREE-----LRHTPELPIAPRLKEGSITQGTPLKYDTGASTTGSKKHVR	1474
Db	1382	-----SEVPKPEDGEPRRHEQFIPIA-----ADKRPPKFE-----RKFD--	1416
Qy	1475	SLIGSPQRTPP-----PVHP--LDVMADARALACYEESLKSRRPGTASSGGSGIARGA--	1526
Db	1417	-----FARERPRQRPTRPQDQKPRFRRLRER--EAASKSEVVAVPTNGTVNNVAQE	1469
Qy	1527	PV-----IYPELGKPRQSPLTIVED-----HGAPPAGHLPRG-----SPVTM	1562
Db	1470	PVNTLGDISGNKTPDLSQNQNSQOANEWTASESSDFENRRERDEKKNADLNAQTVVKV	1529
Qy	1563	RE-----PTPRLQEGSLSSSKA--SQRKLTSTPREIAKSPHSTVPHEHHPHIS-----P	1610
Db	1530	GENVLPPKREIAKESFSSQRPVDRQNRGNGGPPKSGRNPSPGRNRRSGPPSKSGKRG	1589
Qy	1611	YEHLLRGVSGVDLYRSHIPIAFPTSI:PROGIPLDAAAYYLPHRLAPNPTYPHLYPPYLI	1670
Db	1590	FDDQPACTTGVDL-----INGSSAHQO-----	1611
Qy	1671	RGVPDPTAALNRQTIINDYITISQOMHNTATAMAQRADMLRG-LSPRESSLAL-----N	1723
Db	1612	EGVPNGTGGQNKSD-----STGKKREDPKGPKKPEKVDALUSQFDLNN	1655
Qy	1724	YAAPRGRIIDLSQVPHPLVLPVPTPGTPTAMDRLAYLPTAPOFFSSRSHSSSPSLSPG	1783
Db	1656	YAS-----VVIIDHPVTVIEDP-----QSNLNDGGFT	1684
Qy	1784	HLTKPTTTSSERDRDRDRDRDRREKKSILSTVTVEHAPTIWRGTGEOSSGSSGSGG	1843
Db	1685	EV-----VSKQOQKRLQDEER--RKKEEQVI-----QVWNKKNAKEKGRS-----	1722
Qy	1844	GGGSSSRPASHAHQHSPI:SPR-----TQDALQORPSVLHNTGMKGIITAVEPSKPTVLR	1899
Db	1723	-----QTSKULPPRFPAKQATGIIQOAOQO---SASVPPPLASAPLPPSTASV	1764
Qy	1900	STSTSPVRPAANTPPPATHCPGLGTLGDGYVP-TLMEPVLPLKPEAPRVARPERPRADTGH	1958
Db	1765	PASTSAPL-----PATLFPVASTSAPVPASTLAPVLASTSAPVPASPLAPVSASASV	1817
Qy	1959	FLAKPPARS--GLEPPASSKSGSERPLVPVSGHATI-----ARTAKN	2001
Db	1818	SASVPASTSAAAITSSAPASAPAPTILASVSTPASVITILASASIPILASALASTSAPT	1877
Qy	2002	LAPHASDPDP--APASASDPHREKTSQKPSIQI:QELERLSLYGHGSSYSYSGEVPEVSP	2058
Db	1878	PAPAAASPAAPVITAPTIPASAP--TASVPLA-----PASASAPAPAP-TP	1920
Qy	2059	VSSPSLTHDKGLPKHLEELDKSHLEGELRPQGPVVLKGCGEAAHLPHLRPLPESQPSSS-	2117
Db	1921	VSAFN-----PAPAPACTQAO--THKPQONFLOTTSSQSKQ	1955
Qy	2118	-----PLIQTAPGVKGHORVVTLAQHSIVITQDYTRHHPOO--LSAPL-----P	2160
Db	1956	PPPSIRLPSAQTPNG-----TDYVASGKSIQTPQSHGFTLTAELEDWKNKVAPP	2001
Qy	2161	APLYSFFGASCAPVLDLRPPSDLYLPDPDHGAPAR-----GSPHSEGGKRSPEPNKTSV	2214
Db	2002	AVLNDISKULGPI-----SPQO-----PPSVSANNKPLTSPGSAFSSSGAKNGQE---	2048
Qy	2215	LGGEDGIEFVSPPEGMTEPHCHRSASVYVLLYRDGEQ--TEPSRMGSKSPGNTSQPPAFF	2272

Db	2049	LEIGTTIOGAPASNGNE-----NEWVPVLSEKSADKIPEPKEQKQP---RAGPIKA 2100
Qy	2273	SKLTESNSAMVSKKO-EINKLNTNHRN-----EPEYN-----2305
Db	2101	OKLPDLSPVENKEHKPGPKCKESLKNRKVKDAQQVEPEGOEKPSPATVRSIDPVTTKET 2160
Qy	2306	--ISQFGEIIFNNPAITGTGLMTYRSQAQOEHAFTNMGLEAIIRKALMGKYDQWESPP 2363
Db	2161	KAVSEMSTEIGTMI.SVSSAEYGTNAKESVTDYTTTSSSLPNTVATNNTKMEDTLVNNVPL 2220
Qy	2364	SANAFNP-----LNASASLPAAPIITAADGRSDHILTSPGGCGKAKUSGRPSSRKAKSP-- 2417
Db	2221	PNTLPLPKRETIQQSSSLTSVPPT-----FSLTFKMSARKAWENSPNVREKGPSVT 2273
Qy	2418	-----APGLASGDRPPSVSSVHSEGDGCR-RTPLNTRVWEDRPSGAGSTPPFVNPLMR 2470
Db	2274	STAPPIATGSSSASGPFSTANYNSFSSASMPQIPVASVTPFASLSGAGT-----YTTSSLS 2329
Qy	2471	LQAGVNASPP-----PPGLPAGSGPLA 2492
Db	2330	TKSTTTSDPNICKVKPQQLQTSLSPSA 2357
RESULT 42		
ABB71160	ID	ABB71160 standard; protein; 5560 AA.
XX	AC	ABB71160;
XX	XX	26-MAR-2002 (first entry)
XX	XX	Drosophila melanogaster polypeptide SEQ ID NO 40272.
DE	DE	Drosophila; developmental biology; cell signalling; insecticide;
KW	KW	pharmaceutical.
XX	XX	Drosophila melanogaster.
OS	OS	WO200171042-A2.
PN	PN	27-SEP-2001.
PD	PD	23-MAR-2001; 2001WO-US009231.
PF	PF	23-MAR-2000; 2000US-0191637P.
XX	XX	11-JUL-2000; 2000US-00614150.
PR	PR	(PEKE) PE CORP NY.
PA	PA	Venter JC, Adams M, Li PWD, Myers EW;
XX	XX	WPI; 2001-656860/75.
DR	DR	N-PSDB; ABL15263.
XX	XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	PT	genes from Drosophila and for elucidating cell signaling and cell-cell
PT	PT	interactions.
XX	XX	Disclosure; SEQ ID NO 40272; 21pp + Sequence Listing; English.
PS	PS	The invention relates to an isolated nucleic acid detection reagent
XX	XX	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	CC	useful in developmental biology and in elucidating cell signalling and
CC	CC	cell-cell interactions in higher eukaryotes for the development of
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	CC	sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC	CC	ABB72072). The sequence data for this patent did not form part of the
CC	CC	printed specification, but was obtained in electronic format directly
CC	CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	XX	Sequence 5560 AA;
SQ	SQ	

Query Match 3.5%; Score 466.5; DB 4; Length 5560;
Best Local Similarity 18.4%; Pred. No. 1.9e-16;
Matches 600; Conservative 373; Mismatches 1159; Indels 1125; Gaps 142;

QY 31 ARTHDVGLLEYOHSDVASHLSPSIIQPORRRFSLISEFOPGNERSOELHLRPSHS 90
DB 2358 AATNTSGKLHGHRRSVERKSRG-----SDEG-----HS 2390
QY 91 YLPGLCKSEMEFTESKRPRLELPDPLLRPSPLATGQAPAGSBDLTKDRSLTKLEPVSP 150
DB 2391 -----SSKSLRAKLM-----MLSSADSDTDADASKHSIF----- 2420
QY 151 PSPHTDPELEVPPLPSKEELIQNMDRVDREITWVEQQISKKQKQQLLEBAKPPPEP 210
DB 2421 -DIPDCPNVMDYK--VKARSCKNQRAEE-----KKIKAFSOLKQSRK----- 2465
QY 211 EKVPSPPTIESKURSLVQIYD-----ENRKAEEAAHRLGLEGPVELP----- 255
DB 2466 -----KKRS-----TSYDGDSDTEFEDRQHRNSGSSPHGRYPGLSSDDDDDEET 2511
QY 256 -----LYNPSDT-----ROYHENIK-----INQAMRKLI--LY 283
DB 2512 HQRRISSDAERHGGQDQAGASTLADNRVQMQLRLRLCDGDSSEDEIRNVKHS 2571
QY 284 FKRRNH-----ARKQW-----KQKFCORYDQLM 307
DB 2572 FGKNSNSTRIADSESQAPDLITKQBHTAPAEIKREQLSDEEQKFSEHDSNS 2631
QY 308 ALKKVVERIEN-----NPRRAKESKVRYYEKQFPEIKQ--REIQERMQSRVQR 357
DB 2632 IEERKLTEREIKTELGDFFNSSEYTVTKLKEY-----SPETKHKHKKRRLKS----- 2682
QY 358 GSGLSMSAARSE-----HEVSEI-----IDGL-----SE 381
DB 2683 SSTADTSAAGTPLVMTPLTSPISFDVHSSSECKTFONFDLTKTECSSIPLISAGERRKH 2742
QY 382 QENLEKQROL-----AVIP--PMLYDADQRI-----KFIN 411
DB 2743 KERKEKKREKLRNMTATVPNSPTTNDTSEKLSKEERHLKSKSKSDMNSCNTKIYN 2802
QY 412 MNG-----LMADPMKVYKORQ-----VNMWMSOEKETFREKFMQHPNFG 452
DB 2803 SSGAHPSTSPSLPATPTSPASTAQTAKRGEDKMEFIFGIISDEESQFPE--QAETNKD 2859
QY 453 LIASFLEK--TVAECVLYYV-----LTKKNENYKSLVRSYRRRGKSOQQOQQOQQOQQ 507
DB 2860 IIPSSVSTTGPVISAALQTYKQEPSTPNKNEAHIQLTV-----HEPEQQQLERS 2912
QY 508 QQQPMPRSSQBEKEKEKEKEKEKEPEVE-----NDKEDLLKEDTDTSGEDNDEKE 562
DB 2913 LSGSSSSSHADRHREKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 2970
QY 563 AVASKGRKTANSQGR--KGRITSMANEANSEBAITPQOQSAELASWELNNESSRWTEEE-- 619
DB 2971 -----AGRALEAQLMSDFDTKPISEBA--TPSTAATYRS--DMTDFRFSDEON 3016
QY 620 --METAKGLLEHGRNWSAARMVSKTVSOCKNFYFNKRONLDEILLOHKLKMEKER 677
DB 3017 NSVDMTKQGV-----KSEQQ-----BOHKSOKKKK 3042
QY 678 NARRKKKK-----APAAASEEAPPPVVEDEEAGVSGVNEEEMVEEAEALH 725
DB 3043 KKSKEEKQELKLOQQRRESLNPVASTSSA--PTPGKLTVVNQAAKSHAD--LQLDKX 3098
QY 726 ASGNEVPRGCSGPATVNNSSDTESTIPSPHTEAAKTQNGKPKPPATLGADGPPPPPTP 785
DB 3099 ISSPPVKPSPLCLIGDDDD--DALHTPKAK-----PTTPSSRGNDGLTPS--REK 3147
QY 786 PRTRAPIEPTPASEAGTAPTPPAPPS-----PSAPPVVPKEEKEETAAPVVEG 840
DB 3148 PRLISPIPKPTTIANGSTLTQSAETPVSSGTVISSALATTTTSTTAAGVSAAP-----G 3203

QY 841 BEQKPPAAELAVDTGKABEPVKSECTERAEBEGPANGKDAEAAEATAEGALKAEKKEGGS 900
DB 3204 LONSPTSA-----SAQC--KKKESFIFGFDQLDDRRISAVOSISAEFNS 3247
QY 901 -----GRATTAKSSGAPOSDSSATCSADEVDEA----- 929
DB 3248 TSLLDNIADIEPKIPVASPPRAT--KPLDKLEESKSRVTISOETESAVSALLGESFGTSS 3305
QY 930 -----EGDKNRLLSPRPSLLTPTGDPNRANASPOKPLDLKOLKORAAAIIPIQVTKVHE 983
DB 3306 TTDYSLDGWDE--MSSVNELETFT--LVIAEPDEEAALAKAIETAGEP--ASILEE 3356
QY 984 PREDAAPTKAPPAPPPPPONLOPE-----SDAQO-- 1013
DB 3357 PEME--PEREAPDPDPEAEIESEPVVEVLDPEELNKAVQSLKHEDMDIKADTQOSER 3413
QY 1014 -----OPGSSPRGKSRSPAPPADKEAPAAEAQKLPD 1045
DB 3414 DLQIDTDTENDEADSSGSLKIDETVOSSSPE--KSISNNSPTPRETANIDI PNVESQ 3472
QY 1046 PCWTSGLPFP-----VPPREVIKASPHADPSAFSAFPPG--HPLPLGLHDTARPVLP 1099
DB 3473 PKLSNESTPQPSVITKLPFLDTPKTPVAGLPSPVKIEPTISKLOQLVQPVQTVLPAP 3532
QY 1100 -----PTIS--NPPPLISSAKH 1114
DB 3533 HSTGSGISANSVINLDSNVISSCSNTSASATASASISFGSPTASQNAQMPQASTPKQ 3592
QY 1115 PSVLEROI-----GAISOQMSVQL-----HV-----PYSEHAKAPGVPTWGLPLPMD 1157
DB 3593 GPITPOQAIARTOSLIMOPPTISIPEQTFFAVFQPMVLSPOSHHPQOP--GYMYGIRAPSP 3651
QY 1158 PKKL--APFSGVQKE-----OLSPRGAQGPESLCVPTAQEASVLRGALTALGSPVGGSTKGI 1212
DB 3652 HSPHSPFGRGVAGRLVGLQSPVGR--PMVQSPSPQOQVQOQOQOQOQOQOQOQOQOQ 3709
QY 1213 --PSTRV-----PSDAITYRGS-----ITHGTADV 1237
DB 3710 ASPTTRVSSNSPTTSKVNSYQPRNQVQOPQSPKSAEVAEVTTPQLMTIPLQKMTPIQV 3769
QY 1238 LYKGTITRIIGEDSPSLDRGREDSLPKGHVIEYEGKKGHVLSEY----- 1282
DB 3770 PHPTTIKSVTVVQOQATQSOVASSPPLGSLPPHKNVHLNAHQNOQOQOQOQOQOQOQOQOQ 3829
QY 1283 -----GMSVTQCKSKEDGRSGSPHETAAPKRTYDMMEGRVGRATISSA 1325
DB 3830 QQHMQQFMHQMLQROQHMQOQOQLHQSQQIITSAPQHMHQHQOQOQOQOQOQOQOQOQOQ 3878
QY 1326 SIEGLMGRAPPERHSPHHLKEQHHRGSIQTQIGIPRSYV----- 1364
DB 3879 -----QHNQOHLNQLHAQOQHTQKQHOQAQOQFNQOIQOQHOSQOQHQVQOQN 3926
QY 1365 EAQEDYLREAKLLKKEGTTPPPPSRDLTEAVKTQAL--GPLKLKPAHEGLVATVKEAGR 1423
DB 3927 OAOQHLSSQOQ 3982
QY 1424 SIHEIPREELRHTPELPL--APRLKEGSIQTCTPLKYDTGASTT----- 1466
DB 3983 STSIFASQO--HNSQLPARGVPOQHPQQLSHSSPCKPNTLVSVNQGVQVPAILTRVGSH 4040
QY 1467 -----GSKKHVRSILIGSPGRTPFPVPHPLDVNADARAL--ERACYEESLUKR 1511
DB 4041 SQPNQOQQLPHQOQSSSGHPHQKQLSSFGANLPLQTLPLNVIONTPKIIVQOQHIVANQVPP 4100
QY 1512 PGTASSGSGSIARGAPVIVPE--LGKPRQSPLTVDHGAFFAGHLPRGSPVTMEEPTRLQ 1570
DB 4101 PQT-----QGNAIHYPNQOQKOSTTP-----GHV-----EPTP-- 4128
QY 1571 EGSLSSSKASODRKLSTPREI-----AKSPHSTVPEHHPHPIIS--PYEHLIRGVS-- 1619
DB 4129 --AMSAQKTESVSVIRTPPTTGLGAVISANTVGSLLTEENLIKISQKODELIEQDSKE 4186
QY 1620 -----GVD--LYRSHIPLAFDPTSPRGIFLDAAAAAYYLPRHLAPNPTYPHLYP 1666

Db 4187 VDSYWSAKENVIDSVIKKLDTPLA--SKDAKRAVEMQAIAPAPIN---PQGNQSMQA 4241
QY 1667 PYLIRGYPTAALENRQTIINDYITISQOMHNTATAMAORA-----DMLR 1711
Db 4242 ETAL---PTTSMVNS---NDHDEDE-----TETRLQPPAKPIPTVGRPPRGSGSAKR 4291
QY 1712 GLSPRES---SLALN-YAAGRGIIDLSQVP-----HLPVLVPPPTGTPA----- 1752
Db 4292 GRQPRGAKVGFPLNSVTAAPGVDSLVPQDNGVQVRLKPVTAIVTRGKGRPPRN 4351
QY 1753 -----TAMDRLAYLPTAQPFSSSHSSPLSPGGPHTLTKPTTSSSERDRDRDRD 1805
Db 4352 LLLQQOOLQQQLIDIRKGMEMVTSATSTPL-----PTPIPTSSVLTAAEK-KARNOALT 4406
QY 1806 RBREREKSLTSTTTVEHAPIWPCPEQSSGSGSGSGSGSSRP-----ASHSHAH 1858
Db 4407 QACEQNVASQVGTGDIYEFHEDGGEKPKTISSVAPSABDQRPRLTLTINKTQPSIK 4466
QY 1859 QHSPISPRQDALQRPSPVLHNTG-----MKGIITAVEPSKPTVLR 1899
Db 4467 NISEMEQTIQQOQQOQSEVISNTDPIDGDNBSBNCNTRKSRRLQEKEDRSTVDDIIVR 4526
QY 1900 STSTSPVPAATFPFPAHCPGGLTLDGVYFTLMEPVLLPKEAPR----- 1944
Db 4527 NTNT-----PTGTGP---HLPKGAQTP-----PRSGRNAQAOKTDAVQIINA 4566
QY 1945 VARPERPR-----ADTGHAPLAKPPARSGL- 1970
Db 4567 VGRPRSKDKRTIGECTANLIEBVTASNATVAASHLA-PPEGAGVESHVPDQDAKEVEPV 4625
QY 1971 ----PASSPSKGEPRPLVPVPSVGHATIAARTAKNLAPHAGPDPPAPPASADPHREKT 2026
Db 4626 SVVTPISTTAPSVAAVTVFVP-----AMVPVKPTMPOH--PKKALIAAAIESVQAIN 4678
QY 2027 QSKPF-----SIOELELRSLYGHGSSYSPGVEVPVSPVSPSLTHDK-----GL 2070
Db 4679 SSTPSGGLPMHQTAAAPATKQITGGVADAVSKALVDPVTGVTAGMPQEGNLPAAATAA 4738
QY 2071 PKHLELD-----KSHLEGELRPKP--GPVKLGGEAAHLPLRLPESQSPSSPLL 2120
Db 4739 PANSNEDQOAPPPOLOHQOQOQOHPQOQOQANLQINTLIPSLPNTALGKSVQL 4798
QY 2121 QTAPGVKHQVRVTLAQ-HISEVITQDYTR-----HHPQQLSAPL 2159
Db 4799 ETSAAALLNKPSVLKGNASVIOQQQIQIVAPAKPIILQONPLPTVLHQAHTTVRP 4858
QY 2160 PAPLYSFPGASCPVLDLRR-----PPSDLYLPPDPHCAPARGSPHS-----EGGK 2204
Db 4859 POPL-----KARVLNREKNIQOQLTPTKQAVAPQPH-----APHSGHMLLTDTAGNQ 4906
QY 2205 RSPERNKTSVLGGEDGIEPVSPGEGMTSPGHSR-----SAVYPLLVRDGEQTE 2253
Db 4907 QLVQPOILIAHLLHQOQOHLQVNPVPTTAHSPHSRIPISQOQQLGPGASISP-----QOQO 4960
QY 2254 PSRMGSKSPGNTSQPPAFFSKLTESNAMYKSKQBEINKLANTHNEPEYNIQSPGTEI 2313
Db 4961 PQTIVLIKQASAAQPIQL--HVVSRSKASVVPOQOQ---QLPPTSSTGHLQLAKENYSY 5015
QY 2314 F-----NMPATGTGLM--TYRSQAVQEHASTNWGLEAIRKA-----LMGK 2353
Db 5016 APTVLPTLPAVQOQOQOHLQVKNQOQKGAQIQMPHPHGIIMPTHPCGMLLQOKLPAHLQPO 5075
QY 2354 YQOWEESP-----LSANAFPLNASLASLPAAMPITADGRSDHTLTPSGGGKA 2403
Db 5076 QHQLNPSPPGKNPNVLHGLQSQIMPGSVGPPVPSAAVLKTAQOQVNSVVPVAGIRTA 5135
QY 2404 KVSGRPSRRKAKSP---APCLASGDRPPSVSVHSEGD---CNRRT-----PLTN 2447
Db 5136 IPNISPQSQPRVSPVLVPPGISG--VPPPDASLHDLGAVSGRRTQSPPPAHQOQASPIIP 5193
QY 2448 RWEDRPSGAGSTPPFPNPLIMRLQAG-----VNASPPPPGL--PAGSGP---LAGPHAWD 2499

Db 5194 NDSTYRGVTASRDFMLYQHHLMR--GGDYDDKMGSSPPLLELRRPGSGPRTTIAVPH----- 5247
QY 2500 BEPKPLLCQYETLSDS 2516
Db 5248 -----SLOSPQDRTAADS 5260
RESULT 43
ABB60327
ID ABB60327 standard; protein; 3111 AA.
XX AC ABB60327;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 7773.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL04430.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX PS Disclosure; SEQ ID NO 7773; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 3111 AA;

Query Match 3.5%; Score 458.5; DB 4; Length 3111;
Best Local Similarity 18.1%; Pred. No. 2.5e-16;
Matches 534; Conservative 376; Mismatches 1061; Indels 975; Gaps 123;
QY 3 GSTQLVAQVTRATEPRYPHSLSPVQVIARTHTDVGLEYYQHSHSRDYASHLSPG---SII 59
Db 553 GETLL---WRDTPD-----SRRRRD--SLTWKETKADRAAMIREGVDVSV 593
QY 60 QQRRR-----PSLLSEFPQGNERS-----QELHLRPESHVYL 92
Db 594 RTQKLTRFGKSDSSYSDSDSGEQSGTGGGGSTDTSLCDDDDPKSTKSPKQAKL 653
QY 93 PELGKSEMFIEKSRPRLELLPDPLLRPS---PILATGQAPAGSE---DLTKDRSLTGKL 145
Db 654 ARKLKEQKOLAGSRSTSLERQPKSWAPSSHEIPFMLMGTDSGDEKEDSKTEGPIGDO 713

Db 2634 L--RTGENTINDRUSKINS-IKSIDTLCB-EKPYQKEKQRYIDSLFTSLHFAKSSSL 2689
QY 1896 -----TVLRSTSTSSVRPAATEPP-----ATHCPLGGLDGVVPTLMPEVLLPKEAP 1943
Db 2690 EDLSLSRSLSRSGRSIHRSDYAPSRVSEHRSLS-GSADS-----RRSP 2736
QY 1944 RVARPERPRADTGHAFKAK--PPARSGLPFASPSPKSGEPRLVPPVPSVGHATIAARTPAKN 2001
Db 2737 LGNRDTSPLHRRSHRDISRELSPRRLLEEDEERKDRS-----SVVRDN 2783
QY 2002 LAPHPASPPPPAPASAS-----DPHREKTQSK-----PFSIOELELRSIG 2042
Db 2784 LLFNYFADNRSELSSGSLTGFNHKVDROLEETCAKYADRRSACRTPLS-HPYESRTTA 2842
QY 2043 YHGSSYSPEGVPSVPSPLTHD--KGLPKHLELDKSHLEGELRPQGPVKLGGEA 2100
Db 2843 TRHSHTDPVQI-PTNPAGSATATDSPPRPVSPVQRPYPYH-----RSPGGA--GGTP 2892
QY 2101 AHLP-----HLRPLPESQPSSSPLLQATAPGVKGHRV 2132
Db 2893 LYQPGKLEIRHTTVTSTFYDRFLTEKQIERQTHSRP-----PSPSPVSPVPAKSYVEL 2947
QY 2133 VTLAQHISEVITQYTRHHPOQLSAPLAPLYSPGASCPVLDLRRPPSDLYLPPDHGA 2192
Db 2948 CSTSGTSTATSTSTSSSF---MSSSYAGFSFSLPSAS-----NFSYLN- 2989
QY 2193 PARGPSHSGGKSPENKTSVLGGCGDIEPVSPPEGMTEPCHSRSAVYPLLYRDGEOT 2252
Db 2990 ---GS-----GSGSGISLSIP----- 3002
QY 2253 EPSRMGSKPGNTSQPPAFPSKLTESNANVKSKKQEIKNKLNTHNRNPEYNISQPT 2312
Db 3003 ---RASCSDLRSTSGTSTSTSVTTSYV-----PYNFTSSFTS 3040
QY 2313 IFNMPAITGTGLMYRSQVQAEHASTNMGLEAIRKALMGKYDQWESPLGANAPNPLN 2372
Db 3041 RLNDPIITSTAVSTSSLTSTHSTGVYNNPMWSTFLRE-----PLASSISLGSS 3087
QY 2373 ASASLP 2378
Db 3088 ASPLLP 3093

RESULT 44
ID ABB65772
XX ABB65772 standard; protein; 5533 AA.

AC ABB65772;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 24108.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US0009231.
XX 23-MAR-2001; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL09875.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutic and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5533 AA;
Query Match 3.5%; Score 458.5; DB 4; Length 5533;
Best Local Similarity 17.9%; Pred. No. 5.4e-16;
Matches 602; Conservative 370; Mismatches 1047; Indels 1337; Gaps 147;
QY 31 ARTHTDVGLLEYQHHSRDYASHLSPGSIITQORRRPSSLSEFPQGNERSQELHLPESH 90
Db 2358 AATNTSSGKLHHQHRVSVERKSRG-----SDEGH-----HS 2390
QY 91 YLPELGKSEMEFIESKRPRLELLPDLRPSPLATGAPAGEDITKORSITGKLEPVSP 150
Db 2391 -----SSKSLRAKLM-----MLSSADSDTDDASKHSIF----- 2420
QY 151 PSPHTDPELELVPRPESKELIQNMDRVDRREITWVEQIQISKLKKQOOLEEAAKPPPP 210
Db 2421 -DIPDDCNVSMYDK--VKARSKCNMQRAE-----KKIKAFSOLKOSRAK----- 2465
QY 211 EKPVSPPPTIESKHSVLQIYD-----ENRKAAEAHRILEGUGQVQVELP----- 255
Db 2466 -----KRS-----TSYDGDSDTEFEDQHRNSGSSSPHGRYPGLSSDDDDDBET 2511
QY 256 -----LYNQPSDT-----ROYHNIK-----INQAMKKLI--LY 283
Db 2512 HORRISSDAEHGGDNOGASTLADANVRQMOQNRLRLCDGDDSDSEDEIRNVMKHS 2571
QY 284 FKRRNH-----ARKQW-----KQKFCQYDQLMW 307
Db 2572 FGRNNSNTIASDSESQSPAPDITIQQEHPAIPAQEIKREQLSDEBQKFSRHSNNS 2631
QY 308 ALEKKVERIEN-----NPRRAKESKVREYKEOFFEIRKQ-RELQERMQRVGOR 357
Db 2632 IEERKLKTEREIKTELGDYFNSSEYTYTGKLEY-----SPETRKHKHKSRLKS----- 2682
QY 358 SGLSMAARSE-----HEVSEI-----IDGL-----SE 381
Db 2683 SSTADTSAATPLVMTPLTSPIDFVHSSECKTKFDNFDLTKTECSSIPLEISAGERKH 2742
QY 382 QENLEKQML-----AVIP--PMYDADQORI-----KFTN 411
Db 2743 KERKEREKRLNMTTEATVPNSPTNTDSSEKLSKEERHLKSKSKSMNSCNTKIYN 2802
QY 412 MNG-----IMADPMKVYKDRQ-----VMNMSQEKETFREKFMQHPNFG 452
Db 2803 SSGAHPSTSPSLPATPTSPASTAQTQSRGDKWMEFIFGLISDEESQFPE---QAETNKD 2859
QY 453 LIASFLEK-TVAECVLYY-----LTKNENKSLVRSYRRRGSKSQQQQQQQQQQQQ 507
Db 2860 IIPSSVSTTGPVSAALQTYQEPSTPNKNEAHQLTV-----HEPEQQQLERSR 2912
QY 508 QQQPMWRSSQEKDEKEKEAEKEEKEPEVE-----NDKEDLLKEKTDSTSGENDKE 562
Db 2913 LSGGSSSSSHADRRHREKREKREKREKREKREKREKREKREKREKREKREKREKREK 2970
QY 563 AVASKRKTANSQGR--KGRITRSMANBANSEAITPOQSABLASMELNNESSRWTEEE-- 619

Db 2971 -----AGRALEAQLMSDFDTKPISEEA--TPSTAATYRS--DMTDFRFSNEDN 3016
QY 620 --METAKGILLEHGRNWSAIARVMSKTVSQCKNFYFNKGRQNLDEILQOQHKLKMKEKR 677
Db 3017 NSVDWTKQGV-----KSEQ-----EQHKSDDKXXX 3042
QY 678 NARRKKKK-----APAAAGEEAAPPVV-----EDEMEAGVGS- 711
Db 3043 KRSKEEKELIQOORRESLNPVASTSSA-PPTPGKLTNVQAASKHADLQDAXHSS 3101
QY 712 -----GNEEWEVEAEALH-----ASGN-----EVRP--CECSG 738
Db 3102 PPVCKPSPLCLIGDDDD-----DALHTPKAKPTTFSRGNGLTPSRKRELISPIPK 3156
QY 739 PATVNNSS--DTESIPSPTH-----AAKDTGONGPKPATLGCAGPP----- 779
Db 3157 TPTIANSSLTSTQASTPVSSGTVISSSALATPTSTSTAAGVSAARGLDNSPTSASAQCK 3216
QY 780 -----PG-----PPTPPRKT----- 789
Db 3217 KRESFIPGPDGQLDDRISESAVQISAEFNSTSLLDNIADPKIPVASPPRATKPLDKLE 3276
QY 790 ---SRAPIRPTASEATGA-----PTPPAPPSFS 816
Db 3277 ESKSRVTISQEBTESAVSALLGESFGTSTTYSLDGMDMSVNELETPTLVIAEPDEE 3336
QY 817 A-----PPVPVKEEKEETAAAP--PVEEGEEKPPAAEEL-AVDTGKAEEP 862
Db 3337 AALAAKAIETAGEPASILEPEMEPEREAPDPDPEABIESEVPVEVLDPLELNKAVQSL 3396
QY 863 KSECTEABE-EGPAKGDAEAAATAGALKAEKKEGGGRATTAKSSGAPQSDSSATC 921
Db 3397 KHEMDMDIKADTPQSERDLQIDTDTEENPDEADSSGSLKIDETVQSSSSPEKSIENNPS 3456
QY 922 SADE-----VDEAGGDK--NRLLSRPPSLIT-----PTGPRANASPKPLDLK 964
Db 3457 TPRETANIDIPNVSQPKLSNESTPOPSVITKLPLDTPKTVAGLP--PSPVK----- 3508
QY 965 QLKQRAAAIPIQVTKVHEPPREDAAPTAPPA----- 998
Db 3509 -----IEPTISKLOQLVQPVQTVLPAPHSTGSGISANSVINLDLSNVISSCNTS 3560
QY 999 -----PP----- 1000
Db 3561 AASATASASISFGSPTASQAMPQASTPKQGPITPQQAIRTQSLIMQPPITISIEQTP 3620
QY 1001 -----PPONLOPESDAPQPGS-----SPRGKSRSP 1026
Db 3621 HFAPQMWLSPOSHFPQPGTYMVGIRAPSPHSPGRGVAQSRLVGLSPVGRPMVS 3680
QY 1027 APPADKEAFAAEAOKLPGDPCCWTSGLPPVPVPREVIKAS-----PH 1068
Db 3681 QPSPOQVOQTOQOHALITSPOSSNISPLASPTTRVLSSNSPTTSKVNYSQPRNQOVQ 3740
QY 1069 APPDSAFSVAP-----PGHPLPLGLDHTARPLVLPREPTISNPPPLIS 1110
Db 3741 QSPKSAEVAEQTPPLMTIPLQKMTPIQVPHHTIISKVTVVQQAQOSQVASSPPLGS 3800
QY 1111 SAKHPSV-----LERQIGAISQCM-SVOLHVPYSEH 1140
Db 3801 LPPHKNVHLNAHQOQOQVIAKMTAHQHQHQMQFMHQMIQRQHQHQOQOQLHGQSQOI 3860
QY 1141 AKAPVGP-----VTMGLPLMDP----- 1158
Db 3861 TSAPOHMHQHOAQQOQOQHNOHNLNQLHAQOHPYTKOHAQOQOQNOQIOHQSQOQH 3920
QY 1159 -----KKLAPFSGVKQESLSPRGOAGP 1180
Db 3921 QVQOQNAQOQOHLSSQOQHSQOQIQQOHAQOQOQLOQIQKLOQMHGFPQOQKSPQGVG-- 3978
QY 1181 PESLGVPTAQEAASVLRGTALGSPVGGSIITKIGIPSTRVPSDAITYRGSITHGTPA----- 1235

Db 3979 --HLGSGTSIPASQOQNSQLPA-----RGVPOQOHPQ-----QLSHSSPCXKNTL 4021
QY 1236 -----DVLKYGTTIRIGEDSPSRDLDRGREDSLP----- 1264
Db 4022 VSNQGVQPPAILTRVGSHPN-----QOQQLPHQOQSSSGHPHQKQOLSSPGANLPLQTP 4076
QY 1265 -----KGHVIYEGK-----KGHVLSEGGMSVTQCSKEDGSSSGPPHETAA 1306
Db 4077 LNVIQNTPKIIVQOHIIVAQNVPPPPQOTQGNAIHY-----PONQGDSTPFGHVEPT 4127
QY 1307 PKRTYDMEGRV-----GRAISSASIEGLMGRAIPPERHRSPHHLKEQHIRGSIT 1356
Db 4128 PMSAKTSESVSIVRTPTPTTGLAVISANTVGSU-----LTEENLIK--IS 4172
QY 1357 QG-----IPRSVYEAQEDYLREAA-----KLLKREGTPPPPPSRDLTEAYKTOALGPLKL 1407
Db 4173 QPKQDELIEQDSKEVDSDYSAKEVNIDSVIKKLDI---PLASKDAKRAVEMQAIAPAI 4229
QY 1408 -----KPAHEGLVATVKEAGRSIHEIPRELRHTPBLAPRPL-----KEGSI 1451
Db 4230 PNPQCGNOSMAOETALPTTSMNVNSNDHDETEDETROLPPAKPIPTVGRPPGRGSA 4289
QY 1452 TQGTPLKYDTGASTTG-----SKKHVRSILIGSPG-----RTFPVV----- 1487
Db 4290 KRG---RQPRGAKKVGFFPLNSVTAAPGVDSLVVQPGONGVQTRLRKPVTAIVTRGRKG 4346
QY 1488 -HPLDVADARALERACYEESLKSRRPGTASSSGSIARGAPVIVPELG-----KPRQS 1539
Db 4347 RPRNLLLOQOQLOQ---QQLDIQRKGMEMVTSATSTPLPTPTPTSSVLTAEEKARNQ 4403
QY 1540 PTYEDHGAPFAGHLPRGSPVTM-----REPTPRLOEGLSS--SKASQDRKULTSTPREI 1592
Db 4404 ALTQAOEQNVASQVGTQDIYEFHEDGGEPEKPK---TISSVAPSAEDQR---PRLI 4455
QY 1593 AKSPHSTVPEHHP--HPISPYEHLRG-----VSGVDLYRSHIPLAD-----PTSI 1637
Db 4456 L-----TINKTOSIKNISMEQTIQOQOQOQSEVISNTD-----PIGGDNESCNTRK 4504
QY 1638 PRGI-----PLDAAAAYLPRHLAPNPTYPHL-----YPPYLIRGYPTATALENQTI 1685
Db 4505 SRRLOKEQDRSTVDDIIEDVVRNTNPTGTGPHLPKGAQTPPR--RSGRNAQAKKTDAVQ 4562
QY 1686 INDYITTSQMHNTATAMAQRAADMLRGLSPRSSLALNYAAGPRGIIDLSQVPHL----- 1740
Db 4563 IINAVGRPRRSKDKRTIGEQTANLIEVTASNATVAASHLAPPEGAGVESHVQLDAKEV 4622
QY 1741 -----PVLVPTPGTPATMDRLAYLPTAQ-----PFSSRHSS 1774
Db 4623 EPVSVVTPITSTPAPVSAAPVTVVPVPM--VPVKPTMPQHPKKKATAAABIESYQAINSS 4680
QY 1775 SPLSPCGPTHLT-KPTTTSSEERDRDRDREREKSIILTSITTTVEHA--PIWRPG- 1830
Db 4681 IP-SGGLPMHQTAAPATQKITGVAD-----AVSKALVDPVTGVTAGHPQKEGN 4730
QY 1831 --TEQSSGSGSGGGSGSSSPASHSHAHQHSPIPTQDALQQRPSVLHNTGMKIIT 1888
Db 4731 LPAATAAAPANSNEDGQAAPPOLQHQOQOQHPQPPQOQANLQINTLIPSLGNPIT 4790
QY 1889 A-----VEPSK-----PTVLSR 1900
Db 4791 ALGKSVOLETSAAALLNKPVSVLVKGNASQVIOQOQPPQIIVAPAKPIILOQNLPTVLHH 4850
QY 1901 T--STSPVVRP-----AATFPATHCPLGG---TLDGVPYPTIME 1934
Db 4851 AQHTTVRPPQPLKAHVLNREKNIQQOQLPTKQAVAPPPQAHAPSHGMLLTDTAGNQOLVQ 4910
QY 1935 PVLLPKX-----APRVAREPRADTGHAFKAPPARSGLEPASPSSKSGSEPR 1982
Db 4911 PQIARHLQOQHLQVNVPPPTAHSPSPRI-----PSQOQQLGPGASISP-QOQQ 4960
QY 1983 PLVPPVSGHATTARTPAKNLAPHHAS--PDP-----PAPPASADPHREKTSKFSIQELE 2037
Db 4961 POTVVIKQAASAAQOQIHLHVSSKASVVPQOQOQLPPTSTSTGPH-----LQ 5007

QY 2038 LRSGLGYSYSGEGVEVSP-VSSPSLTH-----DKG-----LPKLEELDKSHLEGE 2085
Db 5008 LAKPNY---SYAPTVLTPLPAVQQQQQHLKYQNNQKGAQIQMPH-----CI 5054
QY 2086 LRPKQPGPVKLGGEAAHL-PHLRPLPESQP--SSSPLL-----QTAPGVKGHRVWTIA 2136
Db 5055 IMPTHGMLLQKLPALHLPQHQHLPNPPPGKPNFVLHGLQSGQIMPGSVGSPVPSAA 5114
QY 2137 -----QHISEVITQDTRHHFQQQLSAPLAPLYSPFGASCVPILDRPPSDLYLPPDPH 2190
Db 5115 VLKTAQQQVNVV-----PVAGIRTAIPNIS-PQSOPRVSP--LVLPFGIS 5157
QY 2191 GAPARGSPHSGKSPENKTSVL-----GGEDGTEPVSPPEGMTEPGHS--RSVA 2241
Db 5158 GVPFDPASLUNDSTYRGVTSARFMLYQHLMKGGDYDDKMGSPPLLELRPPGSPRTIA 5217
QY 2242 YPLLVRDGEQTBFSPRMGSKSP-----GNTSQPAFFSKLTES---NSAMVKKKKQKQIN 2291
Db 5218 VP-----HSLQSQDRTAADSQMAQVYVHTRIPPAHFSEIASRGLYDSGALQLEPPPAH 5273
QY 2292 KKLNTNRNEPEVNIHQTEIFNMPAINTGLMTYRSQAVOEHA-----STNWGLEAI 2345
Db 5274 RPTATISVVVPQ-----QMPAVSSGSPFFIGRDSGVQPGSHHHPGKAMDQMDM 5322
QY 2346 IRKALMGK--YDQWESPPLSANAFNPLNASASLPAAMPITAADGRSDHLLTSPG 2399
Db 5323 DRMSIAAVVQQQQLHP-----ALPAGMEL--ASQAPPAMAPP 5363

RESULT 45

ABP69375
ID ABP69375 standard; protein; 1708 AA.
XX AC ABP69375;
XX DT 20-JAN-2003 (first entry)
XX DE Human polypeptide SEQ ID NO 1422.
XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
cell-proliferative disorder; neurodegenerative disease; bacterial;
Parkinson's disease; Alzheimer's disease; autoimmune disease;
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
antiarthritic.
OS Homo sapiens.
XX WO200270539-A2.
XX PD 12-SEP-2002.
XX PF 05-MAR-2002; 2002WO-US005095.
XX PR 05-MAR-2001; 2001US-00799451.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Zhou P, Goodrich RW, Auandi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
DR N-PSDB; AB211592.
XX New polynucleotides comprising sequences assembled from expressed
sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.

PS Claim 9; SEQ ID NO 1422; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences (AB211119-
CC AB212086) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIFO at ftp.wifo.int/pub/published_pct_sequences
XX
SQ Sequence 1708 AA;

Query Match 3.5%; Score 458; DB 5; Length 1708;
Best Local Similarity 20.0%; Pred. No. 1.2e-16;
Matches 408; Conservative 223; Mismatches 673; Indels 736; Gaps 102;
QY 666 LQOHKLMKERNARRKKKAPAAASEAAPPVVEDEMEASGVSGNEEMVEAEALH 725
Db 1 MMHYKLTGTTLETLPLRKRKEKSLGDEE--PPAFESTYKQFGS-----42
QY 726 ASGNEVPRGECGPATVNNSDTESIPSPHTPAKADTQONGPKPPATLGADGPPGPPPTP 785
Db 43 -----PGSDAARNLPLESTKSP-APPSKSV-----PSLEGTGQPRTPKP 83
QY 786 -----PRTSR--APIEPTPASEATGA-----PTPPAP-----812
Db 84 GSGSESGKERRTTSKEISVIQHTSFKSDSLQPSGLEGEDKPLAQFPSPPPAPHGRSA 143
QY 813 -----PPSPAPPVVPKKEEETAAAPPVEEGEQQPPAAEALAVDTGAEPEV 862
Db 144 HSLQPKLVQPNIQVPEILVTEPRDPTPEPPPEKPEK-----TESFQWQ 191
QY 863 KSECTEE--AEEGPAGKD---AEAAEATAEALKAKEKGGSGRATTAKSGAQDSDS 917
Db 192 RSOTLAQLPAELPKPKKRLRLAEMAQSSGESSFESSVP---LPRSPQESNVLSGSR 248
QY 918 SATCSADEVDENAGDKKRLLSRP-----SLTTPGDP-----RANASQKP-----LDL 963
Db 249 SASFERDDHGKAEAPSPSSDMRPKPLGTHMLTVPSHHPHAREMRRSASEQSPNVSHAHM 308
QY 964 KOLKORAAAIPPIQVTKVHEPPREDAAATKPAAPP-----PPQNLQ-----P 1007
Db 309 TETRSKSFYDGLSLT-----GFSAPAPVAPPARVAPERRKCFVLVQASLSRPP 358
QY 1008 ESD---AP--QQPGSSPRGKSRSPAPPADKEAFAAQAQLPGDPPCWTSGLPFP-----V 1057
Db 359 ESELEVAPKGRQSEEPQPSKSAKSLQISSAATSHGPGPGKPGQDRPLGPTV 418
QY 1058 PPREVIKA-----SPHADP-SAFSYA-----PPGH-----1082
Db 419 PYTEALQVPHHPVAQTPLHEKFPYLPVPSLFSFQHLVQHEPGQSPSEFFTSQMSLLSP 478
QY 1083 -----PLPLGLHDTARPVLPRPPTISNP-----PPLISSAKHPSVLERQI---CA 1124
Db 479 YSMPLPPLSFOA--PPLPLOTVLHPQLHLPOLMHPANIPFRQPPSFLPMPYPTSSA 536
QY 1125 ISQGM-----SVQLHVPYSEHAKAPGVPTVMGLPLMDPKKLAPFSGVKQE 1170
Db 537 LSSGFFPLQSQFALQPLGPDVESHLPQIKTSLAPLATGSAG-----LSPSTEYSSD 587
QY 1171 QLSRPGQAGPESLGVPTAQEASVLRGTALGVSFGSITKGP---STRVPSDSAITVR 1226
Db 588 IRLPP---VAPPASSAPTSAPP-----LALPACPDWTWVSVVVRVQTNMPSYGSAMT 639

1227 G-----SITHGTPADVLY-----KGTITRIIGED-----SPSRLDRGREDSLPKG 1266
Db TLSQLVTOQSGSATVALPKFEEPPSKG--TTVCAGADVHEVGGPSGSEEOSAPFPTP 697
1267 H-----VIYEKKGHVLISYEGGMSVTCQSKEDGRSSGPPHETAAPKRTYDM-----MEGR 1317
Db YLRVPVTLPERKKTSLSESILSL-----EGSSSTAGGSKXVLSPAGSLELTWETQOQR 752
1318 VGRAISSASIEGLMGRAIPP-----ERHSPHLLKQEHHRGSIHQGIPRSYVEA 1366
Db VKEEASAKDEKL--ELVCKSVVLSTEDGKPEKSHL-----GNQOQG-----795
1367 QEDYLREAKLLKREGTTP-----PPPSRDLL-----TEAYK--TQALGPLKLPKH 1411
Db RELEMLSSLSDDPDTKEIPLPHPALSHGTAPGSEALKYPOPSG-----KPHR 846
1412 EGLVATVKEAGRSIHEIPRELRHTTPELP--LAP--RPLKEGSIQTCTPLKYDTGASTTG 1467
Db RGLTPL-----SVKEDSKQEDPLSLAPSSILPSE---TSRPAKSEQE---TD 891
1468 SKKHVRSVLIGSPGRTFPVPH-----PLDVM-ADARALERA--CYEBSLKS 1510
Db VLQFSLHTTTNVSWCYLVNPKPHIQHADRRSSVYAGWCISLYNPN 941
1511 RPTASSGSGSIARGAPVIVPELGKPRQSPLTIEDHGAPAGHLPGRGSPVTMREPTRLQ 1570
Db LRGVSTKAALSLR-----SKOKVSKETYTMATAP---HPEAGRLVPSSSSRKPRMT 989
1571 E---GSLSSSKAQDKRLTSTPREIAKSPHSTVP---EHHPHIPISYVHLLRGVSGVDLY 1624
Db EVHPLSVLSEGGQDLARVKEBERGEPEDAPASQORGEPAKIKFE-----GG---Y 1040
1625 RSHIPLAF-----DPTSIPRGIPLDAAAAYILPRHL-----APNPT 1660
Db KSNEEYVYVGRGKGVVCEGIRCKKPSMLKXKHITHDVRVYCKCHFAFKYKGNL 1100
1661 YPHLYPPYLRGVPDTAALENRQTIINDYITSOQHNTATATAQADMLRGLSPRESSL 1720
Db TKHMSKAHSSKQCTGVLE-----ELEAEECTS---DOLFQDSEGREGE 1143
1721 ALNYAAGPRGIIDLQVPHLPVLVPTPGTATAMDRLAYLPTAPQPFSSRHSSSLSPG 1780
Db AV-----BEHQFSDL---1153
1781 GPHTLTKPTTSSSERDRDRDREREKSIILTSTTT-----VEHAPIWR 1828
Db EDSDSLDLDEDEDEESQDELSPSEAPPGPHPALRADSSPILG 1202
1829 P-----GTEQSSGSGS-----SGGGGSSSRPA-----SHSHAHQ 1859
Db POPPDAPASGTEATRGSSVSEABRLTASSCMSQMPGLPWLGPAPLAGSVEKDTGSALS 1262
1860 HSPISPR-----TQDALQORPSVLNHTGMKGIITAVEPSKPTVLRSTSTSPV-- 1907
Db YKVPSPRRPWSKESGRPLARKHSLTKNDS-----SPORCSPAKEPQASAPSP 1314
1908 -----RPAATFPFATHCPGL-----GTLDCGVYPTLMEPVL 1937
Db GLHVDPCRGMGPLPCGSPRLQLSPLTCLGRELAPRAHVLSKLEGTDD--PG-----1365
1938 LPKEAP-----RVARPERPRADTGHAFLAKPPARSGLEPASPSPKSGE-----PRP 1983
Db LPRYSPTRRSPQAESPPR-----SAPPKQWALAGPSGAGEHGPGLGLAPRV 1415
1984 LVPPVS--GHATARTAKNLAHPHA--SPDPPAPASADPHREKTSKPSIQELRLSL 2041
Db LFPFAPLPHLLSRSPETCASPKWASRSRSPSGPA-----HPLSRPFS-----ALHDF 1466
2042 GYHGSSVSPGVPSPVSPSLTHDKGLPKLEELDCKHLEGLRKPQGP--VKLGGEA 2100
Db HGHILALTEINI-----FSH---LPLHQHLTRA-----PCPLIPIGG-- 1501
2101 AHLPLRLPSPESQSSSPLLOTAP-----GVKGHQRVVTVAQHSIVITQDYTRHHPOQ 2154

1502 -----IQVQARPAGAHPTLLPGPTAAWVSGSGGSDLTGARE-----AQERGRWSPT 1550
QY LGAPLPAPLYSFFGASCPVLDLRRPPSDLYLPPDPH-----GAPARGSPHSEGG 2203
Db ASVSPVAKVSKFTLSSELEGDYPKERERTGGGPRPDWTPHGTGA 1600
1551 SSS-----ASVSPVAKVSKFTLSSELEGDYPKERERTGGGPRPDWTPHGTGA 1600
QY KRSPENKTSVLGGSDGIEPVPPEGMTEPHGHSRAVYLLYRDGEQTEPSRMGSKSPG 2263
Db PAEPTPTHS-----PCTPPDTLPRPPOGR-----RAAQSWSPRLSPRAPA 1641
QY N---TSQPAFPSK-----LTESNSAMVSKKQKINKLANTH-----NNNEEYNISQP 2309
Db NPESATPPDLDRSSVGCGLAEA--SARFPARTNLGSEPRTRQDSPKSGSGEPRAPHQP 1700
RESULT 46
AAB41231
ID AAB41231 standard; protein; 2971 AA.
XX AAB41231;
AC AC
XX AC
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF995 polypeptide sequence SEQ ID NO:1990.
DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
XX OS
XX WO200058473-A2.
PN 05-OCT-2000.
XX 31-MAR-2000; 2000NO-US008621.
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
DR N-PSDB; AAC75440.
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX Claim 11; Page 1510-1517; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;

CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC OREX-associated disorder. The nucleic acids can be used to express OREX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 SQ Sequence 2971 AA;

Query Match 3.5%; Score 456; DB 3; Length 2971;
 Best Local similarity 19.6%; Pred. No. 3.3e-16;
 Matches 596; Conservative 305; Mismatches 1141; Indels 1006; Gaps 134;

QY 71 EFQPGNERSQE---LHLRPESHSLPELKGSEMEFI--ESKPRLELLPD--PLLRRSP 122
 DB 91 DFQPGDEDEDEBETLEVEBQOBNDAEQRREIEULRREGELPLEELRSLPPQLLEG 150
 QY 123 LLAATGQAGSEDLTKDRSLTGKLEVPSPSPPHTDPELEBLVPPR---LSKEELIQNMDRV 179
 DB 151 SSPSQTPSHSDTRDPEGAE---EPQV---LEIKPPSAVTQKNQPHDPED 202
 QY 180 DREITVWQOISKLK---KKQOOLEBAEAKPPEPKVPSPPIESKHSLSVLIIY----- 231
 DB 203 DEEFTANEBAEEDETIAAEQLEGEVDHAMELSLAREGEL-----SMEELQOYAGA 257
 QY 232 -----DENRKKAEAAHRILEGGP-QVELPLYNQPSDTRQYHENIKINQANEKLI 281
 DB 258 YAPGSSGSEDEDEVDANSSDCEPGPVEAEEP---POEDSSSQSDSVE----- 304
 QY 282 LYFKRRNHARKQKQFCORYDQLEALBKVKVERIENNRPRRAKESKVEYVEKQPPETR 341
 DB 305 -----DRSEDEDEHSEEBETSGSSASESESESEDAQSQA 343
 QY 342 KORE-----LOERMQRVQGGSLGMSAARSEHEVSEIIDLSEQENLEKQMRQ 391
 DB 344 DEEEDDDDFGVEYLLARDEBQESADAGSPPTPGTTTLPKKEITDIAAAASLOPKYT 403
 QY 392 LAVIP-----PMLYDADQRIKEI-----NMGIMADPMKVYKDRQVMNN----- 431
 DB 404 LATTQVKTPILLRLGRLREYQHIGDLVMTYKELKGLADEMGLGKTIQTISLAHL 463
 QY 432 -----WS-----EQKETFREKFMQHPKQFGLIASFLERKTVAE----- 465
 DB 464 ACEKGNWGPILLIIVPTSVMLNWEMLKWCPSF---KILTYGAQKERKLRQGWTKEN 519
 QY 466 -----CVLYYLT-----KQENYKSLVRRSYRRRGKSQOQOQOQOQOQOQOQOQOQO 507
 DB 520 AFHVCTISYKLVLDHQAFRRKNRYLIIDEAQINKFKSQWQSLNFNFSORLLLTGT 579
 QY 508 -----QQQPMRSSQBEKEKE-----KEKEAKEKE-----KE- 536
 DB 580 PLQNSLMELWSLMHFLMHPVQSHREFKWFNSPLTGMIEGSGQYNEGLVKRLHKVLRFP 639
 QY 537 ---EVENDEKEDLLKEKTD-----DTSGEDNDEKEAVASKRKTANSQGRR 578
 DB 640 LLRRVKVDVEKQPKKYEHVIRCLSRQRCQCLYDDFQAQTTTKETLAT-----GHF 690
 QY 579 KGRIT-----RSMANBAN-----SEEAITPOQSAELASMEINELSSRWTEEMETAKG 626
 DB 691 MSVINILMQLRKVCNHNPLFDPRVTPSPFTTTCFSTASLVLRATDVHPLQRLDMGRFD 750
 QY 627 LLE-HGRNWSAIARMVGSKTVCCKNFYFNYKKRONLDELQOHLKMKERNAREKKK 685
 DB 751 LIGLEGR-----VSRYEADTFLPRHL-----SRRVLLLE 779

QY 686 APAASEEAAAPPPVVEDEMEASGVSGNEEMVVEAEALHASGNVPRGECGSPATVNS 745
 DB 780 VATADPP-----PPRPKPVKMKVN-----RMLQ-----EVPKQEGRTVVVWNP 818
 QY 746 ---SDTESISPHTE-AAKOTGQNGPKP---PATLGADGPPGPP---TPRRTSRA 792
 DB 819 RAPLGFVVRPPPGPGLSAQPT--PGPVQVLPASLMVSASAPGPPPLIPASPPGVLPL 876
 QY 793 PIEPTPASEATGAPT---PPAP-----PSPAPPVVPKKEEETAAAPVVE 838
 DB 877 PLOPNSGSLPQVLPSPGLVLSCTSRPPTLSLKTTPPAPVRLSP-----APPG 926
 QY 839 EGEEQKPAABELAVDTGKAEPVKSECTEAEDEGPAKODKAAEATAAGALKAEKKG 898
 DB 927 PSSLLKP-----LTVPPGYTFPP----- 944
 QY 899 GSGRATTAKSGAPODSDSSATCSADEVDEAEGGDKNRLSPRSLLTPTGD----- 950
 DB 945 ---AAATTT-----SITTATATTAVPAPTPAPORLILSPDMQARLPDSGEVSIQULA 994
 QY 951 -----PRANASPKKPLDLK-----QLKQ-----RAAAIPPIQVTK-----V 981
 DB 995 SLAQRFPVANAGSKLPTFQIQGNKLTLTGAQVQLAVGQPRPLQMPPTMVNTGVVIV 1054
 QY 982 HEPPREDAAATKPAAPPAPPPQNLQSPESAPQ---OPGSSPR---GKSRSPAP----- 1028
 DB 1055 ROAPRDGLTPVPLAPAPRPPSSGLPAVLNRPRTLTGRLTPTLTGTARAPMTPTLVRP 1114
 QY 1029 -----PADKEAFAAEAQKLPDPCWTSGLPFPV-----PREVIKASHAPDP 1072
 DB 1115 LKLVHSPSPVESASA-----PGAAPL-TISSPLHVPSSLPGPASPMPIPNSSPLASVP 1168
 QY 1073 SAFSVAPPCHPLGLHDTARVLPRPPTISNPPPLISSAKHPSVLIERQIGAISQMSVQ 1132
 DB 1169 STVSVPVLSLSPISVPTTLPAAPASAPLTIPISAPLTVSASGALL----- 1214
 QY 1133 LHVPSYSEHAKAVGVPTMGLPLPMDPKKLAPSGVKQEQOLSPPRQAGPPESLGVTQAQ-- 1190
 DB 1215 -----TSVTPPLAPV---VPAAGPPSLAP--SGA-----SPSASA---LTIGLATAPSL 1256
 QY 1191 EASVLGRTALGVPGGSIKGIPTSRVPSDSAITYRGSITHG-----TP 1234
 DB 1257 SSSQTFGHPHLLAPTSSSHVPGNLSTVAPACSPVLVPASALASFPFSAFNPAPAQASLLAP 1316
 QY 1235 ADVLKGTITRIIGBDSRSRLDRGREDSLPKGHVY---EGKKGHVLSYEGGMSYTCQS 1290
 DB 1317 ASSASQALATPLAPMAAPQATAILAPSPAPPLAPLVPALPGAAPVLAASSQTPVPWAPS 1376
 QY 1291 KEDGRS---SSGP---PHETAAPKRYDMMEGRVGRAI--SSASIEGL-MGRAIPPERHSPH 1343
 DB 1377 STFGTSLASASVPVAPTPVLAPSSITQMLPAPVPSPLSPSPASTQTALAPALAPT----- 1431
 QY 1344 HLKEQHHRGSIITQIGIPRSYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQALG 1403
 DB 1432 -----IGGS-----SPSQTLSLGTGNPGPPPTQTL----- 1458
 QY 1404 PLKLPAPHEGLVATVKEAGRSIHEIPRELRHTPELPLAPR-PLKEGSIQTQGTPLKYDTG 1462
 DB 1459 ---LTPA-SSLNPT---PAQTLAPGPPPLGTQTLTSLAPAPLAPASVPGPAPAHTLT 1511
 QY 1463 ASTTSGSK-----HVRSLIGSPGRTPFPVHPLDVWADAPALERACVYESLKRPGTASS 1518
 DB 1512 APASSASLLAPASVQTLTSPA-----PVTLPAAAOQLALAPASTQSPASQASL 1564
 QY 1519 GCSIARGAPVY-----PELCKPROSLTYEDHGAPFAGHLPGRGSPVTMREPT 1566
 DB 1565 VVSAGAAPLFTVMVSRRLPVSKDEPDTLTLRSGSPSPSPSTATSGGPRRPP-----PP 1620
 QY 1567 PR-----LOE-----GSLSSSSASQDRKLTSTPRTIAK---SPH 1597
 DB 1621 PRSPFVLDLSEKKEKQSRERLERIFQLSEAHGALAPVYGVTEVLDFCTLLPQPVASPIGR 1680
 QY 1598 STVPEH-----HPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSPRGIPLDAAA 1647

RESULT 48

ABB59410
ID ABB59410 standard; protein; 1612 AA.
XX
AC ABB59410;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 5022.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL03513.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 5022; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1612 AA;

Query Match 3.4%; Score 452; DB 4; Length 1612;
Best Local Similarity 21.8%; Pred. No. 2.5e-16;
Matches 419; Conservative 149; Mismatches 748; Indels 604; Gaps 90;

QY 560 EKEAVASKGRKTA-----NSQRRKGRITRSMANEAN---SEEAITPQQAELASMLNE 611
Db 4 ESRAATRGLEASASQCNQRRHR--DRS--ANNINGNAAAATAKGVHRCGAICE 61
QY 612 ---SSRW-----EEMETAKKGLLEHGRNWSAIRMVGSKTVCCKNFYFNKKR 659
Db 62 RWKQNRWSHHPLTALVASHQAHAAAVGLKCGGKS----AVCVGLSVQQC----- 107
QY 660 QNLDEILQOHLKQKERNAR-----RKKKAPAAASEAA-----PPVVEDE 703
Db 108 --VDDLTTGPVPC--ALNTRCVTDGLEICVPKSAEAAPTAAVAKWVTITDSPPVVSSEA 163
QY 704 EMEASGVSGNEEEMVEEAEALHA--SGNEVPRCEGSPA-----TVNNSDTSIPS 753
Db 164 PLVDSTNSGSGVEISTEGAPASAVTPKPSIPVETTOIPLETTQVAVETTQIPL 223
QY 754 PHTAAKD-----TGONGP-KPPATLGADGPP-----PG-----PPTP-- 785

Db 224 ETTQASGETTTAAIEETTTGSEAPLEPESTVPSDDTTPVDSTLAPGWEGTYPTIEPNTPAE 283
QY 786 ---PRRTSRAPIETPASEATGAPPPPPAPSPSPAPPPVVPVKEEKEETATAAPPVEEGE 841
Db 284 DFNAPAGTTLAPGEVDPN-----PIDPNSPLDPNAP-----EESTNEPLVDPLPAD 332
QY 842 EQKPPAAEELAVDTGKAEPVKSECTEEAEEGPAKGDAEAEATAEGALKAEKKEGGSG 901
Db 333 TTTAP---DSPVEGSSAAPCTPADVTTAAPGAPADGSSA-APGSPADGSSAAP---GSP 384
QY 902 RATTAKSSCAPODSSSAT-CSADEVDEAEGDKNRLLSRPSLLTPTGDP--ANASPO 958
Db 385 ADVTTAAGAPADGSSAAPGAPADGSSAAPG-----SPADVTTAAGAPADGSSAAPG 437
QY 959 KPLDLKQLKQRAAAIPIQVTKVHEPPREDAAPTKEAPP-----APPPONL-----Q 1006
Db 438 SPAE-----GSSAAPGAPADVTTAAPGAPADGSSAAPGAPADGSSAAPGADVTTAAPGA 493
QY 1007 PESDAPQPGSSPRGKSRSPAPPADKEAFAEAAQKLPDGPCCWTSGLP-----FPV 1057
Db 494 PADGSSAAPGSPAGSSAAPGAPADVTTA-----PGAPADGSSAAPGAPADGSSAAPG 547
QY 1058 PREVIKASPHAP-----DPSAFSYAPPCHPLPLGLHDTARVLPRTTISNP--P 1106
Db 548 SPADVTTAAGAPADGSSAAPGAPADGSSAAPGSPADV---TTAAPGAPADGSSAAPGAP 604
QY 1107 PLISSAKHPSVLERQIGAISQGMVSQVLPVHPYSEHAKAPVGP--VTMGLP-LPMDPKLLAP 1163
Db 605 ADVTTAAP-----GAPADGSSAAPGSPAGSSAAPGAPADVTTAAPGAPADGSSAAP 656
QY 1164 FSGVKOEQLSPRGQAGPPESLGVPTAQEAASVLRGTLALG-----SVFGGSITTKIPS 1214
Db 657 --GAPADGSS---AAPGSPADVTTAAPGAPADGSSAAPGSPADVTTAAGAPADGSSPA 710
QY 1215 TRVPDSATYRGSIYTHGTPADVLYKGTITRIIGEDSPSRDLRGREDSLPKGHVIEGKK 1274
Db 711 PGAPADGS-----SAAPGSPADV----- 728
QY 1275 GHVLSYEGGMSVTQCSKEDGRSS--GPHE--TAAPKRTYDMWGRVGRAISSASIEGLM 1331
Db 729 -----TTAAGAPADGSSAAPGSPAGSSAAPGAPAD-----VTTAAGAPADG-- 772
QY 1332 GRAIPPERHSPHLKEQHIRGSIITGIPRSYVEAQEDYLRRREALKKEGTGTPPPPSR 1391
Db 773 -----SSAAPGAP-----ADGSSAAPGSPA 792
QY 1392 DUTEAYKTOALG-PLKLKPAHEGLVATVKEAGRSIHEIPRELRHTPELPLAPRPLKEGS 1450
Db 793 DVT---TAAPGAPADGSSAAPGAPADGSSAA-----PGSPADADGSS 831
QY 1451 ITQGTPLKYDT---GASTTGSKKHDSVSLIGSP--GRTEPPVHPLDVMADARALERACYE 1505
Db 832 AAPGSPADVTTAAGAPADGSS-----AAPGSPAGSSAAPGSPADVTTAAGAP-----A 882
QY 1506 ESLKSRGPTASSGSGSIARGAPVIVPELKGKQSPITYEDHGAPFAG-HLPRGSPVTMRE 1564
Db 883 DGSSAAPG-APADGSSAAPGSPADVTTAAGAPADGSSAAPGAPADGSSAAPGSPADVTT 941
QY 1565 PTFRIQEGSLSSKASQDRKLTSTPREIAKSPHSTVPEHHPHPIPSYEHLLRGVSGVDLY 1624
Db 942 AAPGAPADGSSAAPGSPADVTTAAGAPADGSSAA-----PGSPAEG----- 983
QY 1625 RSHIPLA-FDPTSIPIRGIPLDAAAAYLPHRLAPNPTYPHLYPPYLIRGVDPDTAALNRQ 1683
Db 984 SSAAPGAPADVTTAAGAPADGSSA-----AP-----GAP----- 1013
QY 1684 TIINDYITSQOMHNTATATAOARADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVL 1743
Db 1014 -----AD---GSSAAPGSPADVTTAAGAPADGSSA----- 1041
QY 1744 VPFTPGTATAMDRLAYLFTAPQPPSSRHSPP-----LSPGPGTHLTKPTTSSSERER 1798

QY 746 -----SDTESIPSHTE-AAKDTQONGKP---PATLGADGPPPGPP-----TPPRTSRA 792
Db 819 RAPLGPVVRPPPGPELSAQT--PGVPQVLPASLMVMSASPAGPLLPASREPPGVLLP 876
QY 793 PIBPTPASEATGAPT-----PPAP-----PSPSAPPPVVRKXEKEEETAAAPVVE 838
Db 877 PLQNGSGLPOVLPSLVLGULGTSRPTPTLSLKPTPPAPVRLSP-----APPPG 926
QY 839 EGEBEQPPAAEELAVDTGKAEEPVKSECTEABEGPAKGDABEABATAGALKAEKKEG 898
Db 927 SSSLLKP-----LTVPPGYTFPP-----944
QY 899 GSGRAITAKSGAPQSDSATSADVEAEAGDKNRLSPSPSLLTPTGD-----950
Db 945 --AAATTT-----STTTATTTAVPAPTAPQRLILSPDMQARLPSGVISGOLA 994
QY 951 -----PRANASPOKPLDK-----OLAQ-----RAAIIPIQVTK-----V 981
Db 995 SLAORPVANAGGSKPLTFIQGNKULITGQVQRLAVGQRPLOMPTTWNNTGVVKIV 1054
QY 982 HEPPREDAAPTKAPPAPPPQNLQPESDAQ-----OPGSSPR---GKRSRPAP-----1028
Db 1055 ROAPRDLGTFVPLAPAPRPPSSGLPAVLNPRTLTFLGRLPTTLGTARAPMTPTLVRP 1114
QY 1029 -----PADKEAFABEAOKLPGDPPCWTSGLPFPVP-----PREVIKASHPADP 1072
Db 1115 LKLHVSPSPVBSASA-----PGAAPL--TSSPLHVPSLLPGPASSPMPPTNSSLASPV 1168
QY 1073 SAFSYPAGHPPLGLLHDARTVLPVRPTTISNPPPLISSAKHPSVLEROIGALISQMSVQ 1132
Db 1169 SSTVSPLSSSLPISVPTTLPAASAPLTIPIASPLTVSASGALL-----1214
QY 1133 LHVPSYEHAKAPGVPTMTGLPLPMDPKLAPFGVGKQEOQLSPRQAGPPES---LGVPTA 1189
Db 1215 -----TSVTPPLAVVPAAPG-----PSLQPSG-ASPASALTLGLATA 1253
QY 1190 Q---EASVLRGTAUGSPVGGISITKIGIPSTRVPSDSAITYRGSITHG-----1232
Db 1254 PSLSSQTPCHPLLAPTSHVPLGNSTVAPACSPVLVPASALASPPPSAPNPAPAQASL 1313
QY 1233 -TPADVLYKGTIRIIGEOSPSRLDREGDSLPKGHIY-----EGKKGHVLSVEGMSVT 1287
Db 1314 LAPASSAQALATPLAPMAQOTAILAPSPAPPLAPLVLVAPSPGAAPVLASSQTPVPVM 1373
QY 1288 QCSKEGDS--SSGP---PHETAAPKRTYDMGRVGRAI--SSASIEGL-MGRAIPPERH 1340
Db 1374 APSSTGTSLASAPVAPPTVPLAPSTQTMPLPAPVPSPLSPASTQTLALAPALAPT--1431
QY 1341 SPHLKEQHHRIGSITOGIPRSYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQ 1400
Db 1432 -----LGGS-----SPSQTLISLGTGNPQGFPTQTLS-----1458
QY 1401 ALGPLKLKPAHEGLVATVKEAGRSIHEIPEELRHTPELAPR-PLKSGSITQGTPLKY 1459
Db 1459 -----LTPA--SSLVPT---PAQTLSLAPGPPGLGPTQTLAPAPPLAPASVPGPAPHT 1508
QY 1460 DTCASTTGSKK---HDVRSLIGSPGRTPTPVHPDVMADARALACRYEESLKSPPGTA 1515
Db 1509 LTLAPASSASLAPASVQTLTUSPA-----PVTILGPAQAOTLALAPASTQSPASQA 1561
QY 1516 SSSGGSIGARGAPVIV-----PELGKPRQSPLTVEDHGAPFAGHLPKRGSPVTMR 1563
Db 1562 SSLVVSASGAAPLPTVMVSRPLPVSKDEPDTLTLRSGLPPSPSPSTATSFSGPRPRRQP---1617
QY 1564 EPTPR-----LOE-----GSLSSSKASQDKLTSTPRETAK---1594
Db 1618 PPPPRGPFYLDLSLEEKRRQRSLERIRIFOLSEAHGALAPVYGTVELDFCTLPPQVASPI 1677
QY 1595 SPHSTVPEH-----HPHPIPSVPEHLRGVSGDLYRSHIPLAFDPTSTIPRGILD 1644
Db 1678 GPKSPGSPHTFTWYTEAAHRAVLFFQORLDQJUSEIERIFVMP-----PVEAPPP 1729
QY 1645 AAAAYLPRHLAP-----NPTYPHLYPPYLI-----RGYPD-----TA 1677

Db 1730 SLHACHPPWLPAPROAAFOELASELWPRARPLHRIVCNMRTQFPDLRLIYDCGLQTL 1789
QY 1678 ALENRO-----TIINDYITSQ-----QMHNTATAMARADMLRG---1712
Db 1790 AVLLROLKAEHRVLIPTQWTRMLDVLVEQLTYHGHLYLRLDOSTRVEQOALMERFNAD 1849
QY 1713 -----LSPRESSLALNYAAGPRGIIDLQVPHLPVLVPPTPTGTATAMDR-----1757
Db 1850 KRIFCFLSTRGGVGNLTGADTVVFYDSW-----NETMDAQADRCHRIGQTR 1900
QY 1758 -----LAYLPTA-----POPFSSRHS 1773
Db 1901 DVHIVRLISERTVEENILKANOKRMLGDMAIEGNGFTTAYFKQOTIRELFDMPLEEPS 1960
QY 1774 SS-PLSP-----GGPHTLKTPTTSSSERDR-----1800
Db 1961 SSVPSAPEEBETVASKQTHILEQALCRAEDDEDIRAATOQAKAEQVAELAEFNENDGPPA 2020
QY 1801 -----DRERDRDREREKSILTSTTVEH-----1823
Db 2021 GEGBEAGRPGAEDDEEMSRAEQETAAALVEQLTPIERVAMKFLEASLEVSREELKQAEQV 2080
QY 1824 -----APIWR-PGTBQSSSG--SSCGGGSSSRPASHSHAHQ--SPISPR 1867
Db 2081 EAARKOLDQAKEEVFRLPOEBEGPGAGDESSCGTGGTHRRSKAKAPERPGTRYSERL 2140
QY 1868 QDALQORPSVLHNTGMKLIITAVEPSKPTVLASTST---SSPVPAATPPATHCPLOGT 1924
Db 2141 RGARAEQOANTP-----VISAHQ-----TRSTTTPRCSPARERVPAPRPTPAS 2190
QY 1925 LDGVPTLMEPVLLPKAEAPRVARPERPRADTGHAFKAPKPARSGLEPASPSRGSPPRL 1984
Db 2191 APAAPAL---VPVPSVAPVISAPNPITILPVHILPSPPPPSQIIPCCSPA--CTPPPA 2245
QY 1985 VPPVSGHATIARTPAKNLAPHIASP---DPPAPPASAS-----DPHREKTQSRFESI 2033
Db 2246 CTPPPAHTP---PPAQTCLVTSPSPLLGPPSPVISASVTNLPLGLRPEALCAQALASP 2302
QY 2034 QELERLSLGVHSS---YSPGVPEVS---PV-----SSPSLTHDKG-LPKHLE 2075
Db 2303 ESLELASVASSETSSLSLVPKDLLPVAVEILLPVSEKNLSLTPSAFSLTLEAGSIPNGQE 2362
QY 2076 ELDKSHLEGELAPKQPGVKLAGEAAHLP-----HLRPLPESQPSSSPLLOTAPGVKH 2129
Db 2363 QEAPDSAEGLTTLVLP-----EGEELPLCVSESNGLELPPSAASDEPLQBLEADR--2413
QY 2130 QRVVTLAQHISEVITQDYTRHHPOQL-SAPLPAPLYSPFGASCPLDLRRPPSDLYLPPP 2188
Db 2414 -----TSBELTEAKTPTSSPEKPQELVTAEVAAPSTSSATSSP-----2452
QY 2189 DHGAPARGSPHSEGGKRSPEPNKTSVLGGEGDIEBPVSPEGMTPEGHSRSAYVLLYRD 2248
Db 2453 -----EGPSAPRPPR-----RRT 2465
QY 2249 GEOTEPSPRMGSKSPGNTSOPPA--PFSKLTESNSAMVKSKQKINKKLATHNRNEPEYNI 2306
Db 2466 SADVELRGOGTGRPG--QPPGPKVLKPLGRLVTVVEEKELVQRRR-----2509
QY 2307 SQPGTEIFNMPIAITGGLMYRSQAVOEHA STMNGLEAIRKALMKYDQWESSPLS--2364
Db 2510 QORGAASLTVPGVSET-----SASPGSPSV--RSMSPG--ESSPIGGP 2549
QY 2365 -----ANAFNPLNASILPAMP--ITAADGRSDHLLT 2395
Db 2550 CBAAPSSSLTPPQBPFIARRHIELGVGTGGSPENGDGALLAITPPAVKRRRRPKKOR 2609
QY 2396 SPGGGKAKVSGRPSRRKAKS---PAPG-----LASGDRPPSVSVSHSEG 2437
Db 2610 SPADAGRVDEAPSSLTKGTNGADVPVGPETLIVADPVLEQLIFGPPQLGPPQVH---2666
QY 2438 DCNRRTPLTNRVWE---DRPSSAGSTPPFPYNPLIMLRLOAGVMA-----PPPPGLP 2485

Db	2667	---RPNPLLSFVKRRRGRPPKARDLDP-----GTTISAGDGNSESRTPQPPHP 271
Qy	2486	AGSGPLAGPHHAWDEEKPILLC 2507 :
Db	2714	SPLTPL-----PPLLVC 2725
RESULT 50		
AAB50363		
ID	AAB50363	standard; protein; 2972 AA.
XX	AC	AAB50363;
XX	AC	AAB50363;
DT	12-MAR-2001	(first entry)
XX	XX	Human SRCAP.
XX	XX	Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
KW	KW	CAMP regulatory element; CREB binding protein; CBP; ATPase;
KW	KW	transcription activation; DEAD box RNA dependent helicase;
KW	KW	adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
XX	XX	
OS	OS	Homo sapiens.
XX	XX	WO2000073467-A1.
PN	PN	
XX	XX	07-DEC-2000.
PD	PD	
XX	XX	25-MAY-2000; 2000WO-US014719.
PF	PF	
XX	XX	27-MAY-1999; 99US-0136620P.
PR	PR	25-MAY-2000; 2000US-00579181.
XX	XX	
PA	PA	(UYSL-) UNIV SAINT LOUIS.
XX	XX	
PI	PI	Chrivia J, Yaciuk P;
XX	XX	
DR	DR	WPI; 2001-061545/07.
DR	DR	N-PSDB; AAC89860.
XX	XX	
PT	PT	Snf2 related CAMP regulatory element (CREB) binding protein (CBP)
PT	PT	activator protein, capable of co-activating CREB binding protein, useful
PT	PT	for modulating transcription and for affecting viral infection.
XX	XX	
PS	PS	Claim 1; Page 86-94; 103pp; English.
XX	XX	
CC	CC	The present sequence is an Snf2 related CREB (CAMP regulatory element)
CC	CC	binding protein (CBP) activator protein (SRCAP) polypeptide. It has
CC	CC	ATPase activity and is capable of activating transcription. SRCAP
CC	CC	polypeptides are useful for activating transcription in a cell, for
CC	CC	enhancing CREB (CAMP regulatory element) binding protein (CBP)-mediated
CC	CC	activation of transcription in a cell, for treating a patient having a
CC	CC	disease involving a function such as insufficient transcription of a
CC	CC	gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
CC	CC	helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
CC	CC	affected by SRCAP protein. Compounds that modulate SRCAP-function, such
CC	CC	as antibodies, antisense molecules, polynucleotides or ribozymes, are
CC	CC	useful for treating diseases mediated by SRCAP-activated transcription,
CC	CC	for example, infection by adenovirus, hepatitis C virus, human
CC	CC	immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
CC	CC	hepatitis B virus
XX	XX	
SQ	SQ	Sequence 2972 AA;
	Query Match	3.4%; Score 452; DB 4; Length 2972;
	Best Local Similarity	19.0%; Pred. No. 5.5e-16;
	Matches 587; Conservative 296; Mismatches 1107; Indels 1092; Gaps 130	
Qy	71	EFQCNERSQ-----LHLRPESHVLPGLGSEMEFI--ESKRPRLELLPD--PLLRFSP 122
Db	91	DFQOEDDEEDDETIEVEEQGNDAQRAEIELLRREGELPFLLEHLRSLPQLLEGSP 150
Qy	123	LLATGQAGSBDLTKDRLSLTGKLEVPSPSPPHPTDPELELVPPR---LSKELIQNDRV 179

Db 1055 RQAPRDLGTPVPLAPAPRPPSSGLPAVLNPRPTLTGRLPTTLGTARAPMTPTTLVRP 1114
QY 1029 -----PADKEAFAAEAKLPGDPPCWTGLPPVP-----PREVIKASHPADP 1072
Db 1115 LKLVHSPSPSEVSASA-----PGAAPL-TISSPLHVFPSSLPGPASSPMPIPNSSPLASPV 1168
QY 1073 SPSYAPPCHPLPLGLHDTARPVLPRPPTISNPPPLISSAKHPSVLEROIGALISQGMVQ 1132
Db 1169 SSTVSFVLSLSPISVPTTLPAAPASAPLTPIASPLTVSASGALL-----1214
QY 1133 LHVPSYSEHAKAPVCPVTMTGLPMDPKLAPFGSVGQQLSPRGQAGPPES-----LGVPTA 1189
Db 1215 -----TSVTPPLAPVPAAPG-----PSLOPSG-ASPSASALTGLGATA 1253
QY 1190 Q-----EASVLRGTALGVSFGSITKIGIPSTRVPSDAITYRGSITHG-----1232
Db 1254 PSLSSQTPGHPLLLAFTSSHVPLNSTVAPACSPVLVPASALASPPSPAPNAPAAQASL 1313
QY 1233 -TPADVLYKGTITRIIGEDSPSLDRGREDLSLPGHVIY-----EGKGHVLSYEGGMSVT 1287
Db 1314 LAPASSASQALATPLAPMAAPQOTAILAPSPAPPLAPLPLVLAPSPGAAPVLASSQTPVPVM 1373
QY 1288 QCSKEDGRS--SSGP-----PHETAAPKRTYDMMEGRVGRAI--SSASIEGL-MGRAIPPERH 1340
Db 1374 APSTPCTSLASAPVPAPTPVLAPSTQTMWLPAPVSPPLSPASTQTLLALAPALAPT-- 1431
QY 1341 SPHLKEQHIRGSIOTIGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTQ 1400
Db 1432 -----LGGG-----SPSOTLSLGTGNFGPPFTQTLG-----1458
QY 1401 ALGPLKLPKPAHEGLVATVKEAGRSIHEIPREELRHTEPELPLAPR-PLKEGSIOTGTPLYK 1459
Db 1459 -----LTPA--SSLVPT-----PAOTLSLAPGPPGLPTQTLAPAPPLAPASPVGPAPHT 1508
QY 1460 DTGASTTGSKK---HDVRSILGSPGRTFPVPHPLDVMADARALACRYEESLKSPPGTA 1515
Db 1509 LTLAPASSASLAPASVQTLTSPA-----PVPTLGPAAQTLALAPASTQSPASQA 1561
QY 1516 SSSGGGIARGAPVIV-----PELGKPRQSPLTYEDHGAPPAGHLPRGSPVTMR 1563
Db 1562 SSLVVSASGAAPLPTVMVSRPLPVSKDEPDTLRLSGPPSPSTATSEFGPRPRQP-----1617
QY 1564 EPTPR-----LOE-----GSLSSKASQDRKLTSTPREIAK-- 1594
Db 1618 PPPRSPFYLDLSLEEKRKQRORSELERIFOLSEAHGALAPVYGTVELDFCTLPPQVASPI 1677
QY 1595 SPHSTVPEH-----HPHPTSPVEHLRLGVSGVDLYRSHIPLAFDPTSTPRGIPLD 1644
Db 1678 GPRSPGSHPTFTWYTEAHRAVLFPQORLDQLSIEIERFVMP-----PVEAPPP 1729.
QY 1645 AAAAYTLPRHAP-----NPTYPHLYPPYLI-----RGYPD-----TA 1677
Db 1730 SLHACHPPWPLARQAAFOQLASELWPRAPLHRIVCNMRTOFPDLRLQYDCGKLOTL 1789
QY 1678 ALENRQ-----TIINDVITQ-----QMHNNTATAMAQADMRLG-----1712
Db 1790 AVLLROLKAEGRVLIPTOMTRMLDVLQELTYHGHLYRLDGSSTRVQEQALMERFNAD 1849
QY 1713 -----LSPRESSALNAAAGPRGIIDLQVPHLPVLVPPPTCTPATAMDR-----1757
Db 1850 KRIFCILSTRSGGVNLTGADTVVFDSDW-----NPTMDAQADQRCHRTGQTR 1900
QY 1758 -----LAYLPTA-----POPFSSRHS 1773
Db 1901 DVHLYRLISERTVEENILKANKQKMLGDMALIEGNGFTTAYFKQOTIRELFDMPLEPSS 1960
QY 1774 SS-PLSP-----GGPHTLTKPTTSSSERDR-----1800
Db 1961 SSVPSAPEEETVASKQHILEQALCRAEDEDIRAAQKAEQVAELAEFNENDGFPA 2020
QY 1801 -----DRDRDREREKSLTSTTVEH-----1823
Db 2021 GEGEEAGRPGAEDEMSRAEQEIAALVQELTPIERYAMKFLASLEEVSRLEELKQAEQV 2080

QY 1824 -----APIWR-PCTEQSSSGSG--SSGGGGSSSRPASHSHAHQH--SPISPRT 1867
Db 2081 EAARKOLDQAKBEVFRLPQEEBEGPGAGDESSCGTGGTHRRSKAKAPERPGTRVSERL 2140
QY 1868 QDALQORPSPVHLNTHGKGIITAVEPSKPTVLRSTST-----SSPVRPAATFPFATHCPGCGT 1924
Db 2141 RGAETQGANHTP-----VISAHQ-----TRSTTTTPRCSPARERVPRAPRPTPAS 2190
QY 1925 LDGVYPTLMEPVLLPKAPRVARPERPRADTGHAFIAKPPARSGLEBPASSPSKSEPRPL 1984
Db 2191 APAIATL-----VVPVVSAPVPIAPNPITILPHILPSPPPPPQIIPCSSPA--CTPPPA 2245
QY 1985 VPPVSGHATARTPAKNLAPHASP---DPPAPPASAS-----DPHREKTQSKPFSI 2033
Db 2246 CTPPPAHTP---PPAQTCLVTPSSPLLLGPPSPVISASVNLPLGLRPEALCAQALASP 2302
QY 2034 QEILESLGLVHGSS-----YSPEGVEVS-----PV-----SSPSLTHDKG-LPKHLE 2075
Db 2303 ESLELASVASSETSSLSLVPPKDLLPVAVEILPVEKLNLSLTPSASLTLLEAGSIPNGQE 2362
QY 2076 ELDKSHLEGELRPQKPGPKVLGGEAAHLP-----HLRPLPESQPPSSPLLOTAPGVKGH 2129
Db 2363 QEAPDSAGETTLTLP-----EGEELPLCVSENGLELPPSAASDEPLOEPLADR-- 2413
QY 2130 QRVVTLAQHISEVITQDYTRHHHPQOL-SAPLPAPLYSPFGASCPCVLDLRRPPSDLYLPPP 2188
Db 2414 -----TSEELTEAKTPTTSPEKQELVTAEVAAAPTSSSATSSP-----2452
QY 2189 DHGAPARGSPHSGGKRSPENKTSVLGGEDGIBVSPPEGMTEPGHRSASVYLLYRD 2248
Db 2453 -----EGPSPARPPR-----EGPSPARPPR-----RRT 2465
QY 2249 GEOTEPSPMGSKSPGNTSOPPA--PFSKLTESNSAMVSKKQOEINKLNTHNRNEPVNI 2306
Db 2466 SADVEITRGQGTGRG--QPPGPKVLKLPQRLVTVEEKELVQRRR-----2509
QY 2307 SQPGTIFNMPAITGTGLMTYRSQAVQEHASTNMGLEAIRKALMGKYDQWEESSPLS-- 2364
Db 2510 QQRGAASLTLPVGSET-----SASPGSPSV--RMSGP-----ESSPIGGP 2549
QY 2365 -----ANAFNPLNASASLPAAMP--ITAADGRSDHLLT 2395
Db 2550 CEAPSSSLPTPPQPPFIARRHIELGVTGGSPENGDGALLAITPPAVKERRRGRPPKKNR 2609
QY 2396 SPGGGKAKVSGRPPSSRKAKS---PAPG-----LASGDRPPSVSVHSEG 2437
Db 2610 SPADAGRQVDEAPSTLKGTNGADVPQGPETLIVADPVLEPOLIPGQPLGQPVH-- 2666
QY 2438 DCNRRTPLTNRVME---DRPSSAGSTPPFPYNPLIMRLQAGVMA-------PPPPGLP 2485
Db 2667 ---RPNPLLSPEVKERRRGRPPKARDLP--GTISSAGDGNSESRTQPPPH 2713
QY 2486 AGSGPLAGPHHAWDEPKPLLC 2507
Db 2714 SPLTLP-----PPLLVLC 2725

RESULT 51

ABR42219

ID ABR42219 standard; protein; 2193 AA.

XX

AC ABR42219;

XX

DT 28-JUL-2003 (first entry)

XX

DE Human protein kinase 85924.

XX

KW Human; protein kinase; 85924; enzyme; gene therapy; cytostatic.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Modified-site 6. .11 /note= "N-myristoylation site"
FT
FT Modified-site 31. .34 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 35. .38 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 42. .47 /note= "N-myristoylation site"
FT
FT Modified-site 67. .69 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 136. .138 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 143. .148 /note= "N-myristoylation site"
FT
FT Modified-site 154. .157 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 154. .156 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 174. .177 /note= "casein kinase II phosphorylation site"
FT
FT Domain 181. .439 /note= "protein kinase domain"
FT
FT Modified-site 190. .195 /note= "N-myristoylation site"
FT
FT Modified-site 191. .193 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 203. .206 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 215. .218 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT
FT Modified-site 218. .221 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 250. .252 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 252. .255 /note= "Amidation"
FT
FT Modified-site 267. .272 /note= "N-myristoylation site"
FT
FT Modified-site 268. .270 /note= "protein kinase C phosphorylation site"
FT
FT Active-site 305. .317 /note= "serine/threonine protein kinase active site signature"
FT
FT Modified-site 325. .325 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 333. .335 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 335. .338 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT
FT Modified-site 393. .396 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT
FT Modified-site 398. .403 /note= "N-myristoylation site"
FT
FT Modified-site 422. .428 /note= "tyrosine kinase phosphorylation site"
FT
FT Modified-site 454. .457 /note= "Amidation"
FT
FT Modified-site 456. .459 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT
FT Modified-site 492. .495 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 517. .520 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 517. .519 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 600. .603 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 604. .607 /note= "glycosaminoglycan attachment site"
FT
FT
FT Modified-site 605. .610 /note= "N-myristoylation site"
FT
FT Modified-site 625. .628 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 746. .751 /note= "N-myristoylation site"
FT
FT Region 774. .795 /note= "Leucine zipper signature"
FT
FT Modified-site 800. .805 /note= "N-myristoylation site"
FT
FT Modified-site 1064. .1069 /note= "N-myristoylation site"
FT
FT Modified-site 1074. .1079 /note= "N-myristoylation site"
FT
FT Modified-site 1079. .1082 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1079. .1081 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1089. .1094 /note= "N-myristoylation site"
FT
FT Modified-site 1106. .1109 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT
FT Modified-site 1108. .1110 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1113. .1116 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1149. .1151 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1179. .1182 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1188. .1202 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1204. .1209 /note= "N-myristoylation site"
FT
FT Modified-site 1218. .1223 /note= "N-myristoylation site"
FT
FT Modified-site 1221. .1224 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1242. .1245 /note= "Amidation"
FT
FT Modified-site 1242. .1244 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1288. .1291 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1288. .1290 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1332. .1337 /note= "N-myristoylation site"
FT
FT Modified-site 1339. .1342 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1355. .1360 /note= "N-myristoylation site"
FT
FT Modified-site 1362. .1365 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1386. .1391 /note= "N-myristoylation site"
FT
FT Modified-site 1398. .1401 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1398. .1400 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1463. .1466 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1467. .1470 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1482. .1484 /note= "protein kinase C phosphorylation site"
FT
FT Modified-site 1485. .1488 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1508. .1511 /note= "casein kinase II phosphorylation site"
FT
FT Modified-site 1523. .1525 /note= "RGD cell attachment site"
FT

1377 LLKRE-----GTPPPPPPSRD---TEAVKTOALGPL-KLKPAHEG-----LVATVK 1419
1194 MLESDTDADRGSDPGTSPHLSTCGLTGTEESRQANAPVQVNLHKGKWFICPVA 1253
1420 EAGRSIHEIPREELRHTPELPAPRLKESITQGTPLKYDGTGASTGS--KKHDVRSLLI 1477
1254 E-----HPAP-EAPESPPLPLSSLP-PEAS-QDSAPYKQDLSSEKQPSFLASQQLLSQA 1305
1478 G---SPGRTPFPVHPLVDNADARALERACYESLSKRP--GTASSGGGSIARGAPV-IVP 1531
1306 GPSNPFGAPAPLAFESPVTALPQDGAAPATSTMPEPASQASQAGGP---GTQGLTS 1362
1532 ELGKPRQSPLTYEDHGAPEAGH-LPRG-SPVTWREPTPLQEGSLSSSKASQDRKLTSTP 1589
1363 EL---ETSQPLA-ETHEAPLAVQLVGLAPCT---PAPE-----AASTRDASAPREPLPP 1412
1590 REIAKSPHSTVPE---HHPHPTSPYHLLRGVSGVDLYRSHIPLADPTSIIRGPIPLDAA 1646
1413 AP-EPSPHSGTQPALGQAPLPLP-----AAGVAVSLATSQLP-----SPPLG----- 1454
1647 AAYLPRHLAPNPTYPHLYPPYLIR-----CYPDTA--ALEN--RQTIINDYITS 1692
1455 -----PTVPP-QPPSALESDEGPPRPVGFVDTIKSLDEKURTLTYQSHVPT 1501
1693 QQMHNHTATAMAQR---ADMRLGLSPRESSLALNYAAGRGIIDLQVPHLPVLVPPPTG 1749
1502 SSASACTPVEVGDRDFTLEPLRGDQR-----SEVCGDLALPPVPK 1543
1750 TPATMDRLAYLPTAPQPPSSRHSSPLSPGGTHLTKTPTTSSSRER---DRDRERDR 1806
1544 EAVSGRVQL-----PQLVKSSELAFTRGAVMEQGTSSMTSESSPRMLGYDRD----- 1592
1807 DREKXSILTSITVTEHAP-----IWRPG-----TEOSSGSGSS-----GGGSGSS 1848
1593 -----GRQVADSHVPSVQDVPFAPVAPARVEPTDRGGEAGESSAEPSPDMGTVGQA 1648
1849 SRPASHAHQHSFISPTQDALQORPSVLHNTGMKGIIITAVEPSKPTVLRTSTSSPVR 1908
1649 SHPQT-----LGARALGSPKRP-----EQQDVSSPAK 1676
1909 PAATFPATHCPGGTLDGYYPTLMBPVLLPKEA-PRVARPEPRADTGHAFKAPKAPS 1967
1677 TVGRFVSWS-----TQDEW--TLASPHSLRYAPPDVVLDLDEAPSPDVKLAVRRAQTAS 1728
1968 GL-----EPASSPKSGSERPLVP-----PVSGHAT---TARTAKNLAHPHIA---SP 2009
1729 SIEVGGEFVSSSGDEGPARPVPVQKASLPVSGSVAGDFVKATATFLQPSRAGSLGP 1788
2010 DPPA-----PPASADPHREKTSQKPFISIQELRLSLGYHSGSSYSPGVEPVSPVSPS 2063
1789 ETPSRVGMKVPTISVTSFHSQSSYISSDNDSELE-----1822
2064 LTHDKGLPKHLELDKSHLE--CELAPKQPGPKVLGGEAAHPLRLPLP-----2110
1823 ---DADIKKELQSLREKHLKEISELOSQKQEI-----EALYRLGKPLPENGVFFHTAPP 1875
2111 -----ESQSSSPLQATQPVKQHORVTLAQHISEVITQDYTRHHHPQLLSAPLAP 2162
1876 TGRRTKSKLXAGKLLN--FLVRQLKVVASSTGHAD--SSRGPPAKOPAQASVGLTA- 1931
2163 LYSFPGACSVLDLRP-----PSDLYLPDPDHGAPARGS-----PHSE--GGKRRP 2207
1932 --DSTGLSGKAVTQPCSVRASLSSDSCGLASDGGGARGQGWTVVHTSERVTYKSSS 1989
2208 EPNKTSVLGGGEGIEPVSPGEMTEGHSRSVAVPLLYR---DGEOTEPSRMGSKSPGN 2264
1990 KPARFLSG-----PVS-----VSIWALKRLCLGKHSRSTSSSLAPGP 2030
2265 TSOP-PAFFSKLTESNAMSVMKSKQKQKINKLNTHNRNEPEYNISQPGTEIFNPAITGTG 2323
2031 EPQGPQALHVAQVNNNS---NNKKGFTDDLH-----2059
2324 LMYRSQAOEHASTNMGLEAIITRKALMGKYDQWERSPPLSANAFNPLNASLPAAMPI 2383

Db 2060 -----KLVDWTSKTVGAQL--KPTLNQLKQTKQLQMEQAAG-----2096
QY 2384 TRADGRSDHTLTSPPGGGKAKVGRPPSSRKAKSP-----APGLASGDRPPSVSVHSEGD 2438
Db 2097 WAAPGEA-RAMTAPRAG-----VGMPLRPPAPGPLSTTVIPGAA-----PTLSVPTPDGA 2145
QY 2439 CNRRTPLTNRVW 2450
Db 2146 LG--TARRNQW 2155
RESULT 52
ABR41356
ID ABR41356 standard; protein; 2735 AA.
XX ABR41356;
AC ABR41356;
XX 02-JUN-2003 (first entry)
DT Human DITHP transcription factor.
DE Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
XX cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antitense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW transcription factor.
OS Homo sapiens.
XX WO200297031-A2.
XX 05-DEC-2002.
XX 27-MAR-2002; 2002WO-US010056.
XX 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Feralta CH, David MR, Lewis SA, Chen AJ, Panzer SR, Harris B,
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
DR N-PSDB; ACC46296.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 891; 591pp; English.
PS The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP

QY 713 NEEMVEEBAALHASGNEVPRGCS-----GPATVNNSSDTESTIPSPHTEAAKD 761
Db 675 GEPSIAGEEEIIVKVT---TPAGESIAGEEBEIVKVTTPAGESSEGESEEEIKVTTFA--- 728
QY 762 TGONGKPPATLGDGPPGPPPTPP-----RRTSRAPITPTPASBATGAPTPPPAPPS 815
Db 729 -GSSSEGDEEIVKESTPAGEPISGEEDVIAKTTSA-----PKSDLEGVKEPETATE-- 780
QY 816 SAPPVVVPEKEEBEETAAAPVVEGEEKPPAAELAVD--TGKABEPPVKSECTEEABEG 873
Db 781 -----VPABEVEDFAKPTPIABAEBE-PIAGTPTDGTSGE-BEIVKGTTPQTLEEQ 832
QY 874 PAKGDAEAAEATBAGALKAKEGSGRATTAKSCAPQODSSATCSADEVDEABEGD 933
Db 833 P-----EISESTEVPVADDDLSSTASAIASSTEGVQDAASETTSSA-----PARAGD 882
QY 934 KNLLSPRPSLLPTGDPRANASPKPDLKOLKORAAAIPIQV-----TKVHEPPR 986
Db 883 KD-----EATTVPTAQDK-----DDEVEQDADLPLVEDVQOSTTAKTTTTTEQPK 928
QY 987 EDAAPTCKPAPPAPPQNLQPEDSAPOQPGSSPRGKSRSP-APPADKEAFAPAAQKLPD 1045
Db 929 EES-----STEADAEIEVTTSSPADKQEVPEASPADKOHKDEEDVQTATD 974
QY 1046 PPCWTSGLPPVPREVIVKASHPADPSAFS-----YAPP-GHPLPLGL-----HDT 1091
Db 975 LPI-KSDIGPPVVDTEATTQCPETSDETDATDKPPSVLPPVPSQEVPSSTAKVDNRNDFET 1033
QY 1092 ARVLPRPPTISNPPPLISSAKHPSVLERQIGALSQMSVOLHPVSEHAKAPVGPVTMG 1151
Db 1034 EKTLPPSGEDQSEPL-----PAMDLPAGIPGEDCL-----VEGKYANNTIVPAT-- 1081
QY 1152 LPLPMPKPLAPFSGVKQEQLSPRGOAGPPESIGVPTAQEASVLRG---TALGASVPGSI 1208
Db 1082 --APCDVSCKICISLVAQOM---ECKLPENLEKCTV-AADLLDCCPTVICDESTESA 1134
QY 1209 TKGIPSTRVPSDAITVRGSI-THGTPADVLVYKGTTR-----II 1247
Db 1135 EKDEESTAKPDNKIDEDVSEIESTEEIPKQVIMPTGITEQLSHVKPDEEIQPVTSPVAQF 1194
QY 1248 GEDSPRLDR-----GREDSLPGHVIYEGKKGHVLSEGGMSVTOCKSKEDGRSSGPPH 1302
Db 1195 DESTTAKVDKKPIDESAEDKKPIGE-----SEEDSKPIDE-SEEDKK-----PVE 1238
QY 1303 ETAAPKRTYDMBGR---VGRAISSASIE-----GLMGRAIPPERHSPHLLK 1346
Db 1239 ESAEDKKPVEDSEKELPTVIPASIEIEKSEPEDEKKTADFAAPTEQPEATTAQIA 1298
QY 1347 E--QHIRGSI-TQGIPIRSVVEAQEDVLRREAKLLKREGTPPPP-----PPSRDLTBAYKT 1399
Db 1299 DTAEKEVDKLTATTSAPVS---GEDEL-KPADEKKRTETAQIPDAEIPASTDEPESSTE 1353
QY 1400 QALGPLKLKPAHEGLVATVKEAGRSIHEIP-----REELRHTPEL-PLAPRPLKEGSITC 1453
Db 1354 LPTVLDKKPEEDSTKGTEAPESDKVPEVPTSASTENEIESDKFTTVPAPKISASDETE 1413
QY 1454 GT-----PLKYDTGASTTGKCHDVRSILGSPGRTFFPPHPLDMADARALERA 1502
Db 1414 PTAEBEDLVPAFTFPIESEFEVST---KKPAVO---GPPLTLAPAQEKKPVDAETSTE 1467
QY 1503 CYE-----ESLKRPGTASSGSGSIARGA-----PVIIVELGKPROSPLTYEDHG----A 1548
Db 1468 DISTEPSAEVEKASGETSESDEINIDAGASSTVPVPSADEDKTPTSTEKTEADDDKFTTVA 1527
QY 1549 PFAG-----HLPR-----GSPVTWREPTPLQEG----- 1572
Db 1528 PLAGDEESNLKLPQODIFEEEPAPVAVTTAAPSODGEOQKPVVEKEKPIEDGOKPIEDET 1587
QY 1573 -----SLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHIPISYEHLLRGVSGDLYRSH 1627
Db 1588 STTSENIELEPSDRATTIAPSEKPESEFSTCAPTKDEFAEASTDAPESDESEKPESE 1647
QY 1628 IPLAFDPTSIPIRGIPIDAAAAYYLPHRLAPNPTYPHLYPPYLRGYPDTAALENRQTIIN 1687

Db 1648 VP-----TTVAPAGEKIPSTSS---ITPDEEFTATSAVAKP-----DEDEKSTSTIPT 1694
QY 1688 DYTTSQOMHNTAT-----AMAQRADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLP- 1741
Db 1695 DAPASSEEDENSTDOIPEVEPEKKETPAQTPEEGDI-VGATAAPTTSDEVPVQRLPE 1753
QY 1742 -VLVP-PTPG-----TPATAMDRLAYLPTAPQPFSSHRHSSPLSPGPGTHLTKP 1788
Db 1754 EVLAEIPQSTETGKIQQDETTAAPSIDRKE---PYVTIDEDEATTVAPISEKD-----EKP 1807
QY 1789 TTTSSSERERDRDRDRDRERERSKILSTTTTVEHAPIWRPPTQES--SGSSGSGGGGGS 1847
Db 1808 TEEKPVQKPTGEEPESEEEKPIEQDVST--EGPV---STEASEAGSTESS----- 1856
QY 1848 SSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPV 1907
Db 1857 -----EEVKPSTGEVAEKP-----EDKQP-----SSTAQ 1881
QY 1908 RPAATPPP-ATHCPLGGTLDGVVPTLMEPV-----LLPK-----EAPRV--A 1946
Db 1882 APVETIPEISTELF---AQDGDKPTSEAPVDSDETSAPSDEKIPSVSGEEVEPEVTTA 1938
QY 1947 RPERPRADTGHAFLAKPPARSGLPEASS-----PSKGEPRPLVPVPSGHAT 1993
Db 1939 SPOAEEDE-----LKTPAES--EPSSTDKVPETEXQKPEDETKADETPESVTOQSDVAT 1991
QY 1994 IARTPA-----KNLAPHASPDPP-----APPASASDPHEKTQSKPFSIQELE-- 2037
Db 1992 STSAPVAGDIEKDEQATTASPEEEIEKPTIAPAAEIQP-----SEKEPVDSQVESGT 2047
QY 2038 ---LRSGLYHGSSYSGEGVEPVSPVSSPSLTHDKGL-----PKHLEEL----- 2077
Db 2048 KATPAESDQPIDEAPATSGPIDEASTAAPTKEESTTVAASAASPAVHDDDKI KDVTTTQP 2107
QY 2078 --DKSHL-----EGELAPKQPGPVKLGEAAHLPHLAPLPESQP 2114
Db 2108 VADEKEVAAPQDETKTSIDVSTDSPTAQDDEKODKTEAPV-----APTTVSSP 2155
QY 2115 -----SSPLLQTAGVKGHORVV-----TLAQHISEVITQDVTTRHHPOQLSAPL-- 2159
Db 2156 TADSAADSTPTVEVSPVEIDTKPMDDIMSQIIAHTADGAASTEDDEDO---APTV 2212
QY 2160 -----PAPLYS-----PFGASCPLDLRRPPSDLYLPPDPHGAPA 2194
Db 2213 SPQDAEKTVPVAPQSDKTPSSEAPQDADEIPATATPLDDNKIPAT--VAPQTDGCPA 2270
QY 2195 RGSHPHSEGGKRSPEPNKTSVLGGGEDGI-----EPVSPPEGMTFEGHRSASV 2241
Db 2271 TAAPLDE-----DKIQTAAAPLDEEKIPSTAAPLDEKIPAPVSPVDFVPSSEKPAV 2324
QY 2242 YPLLVRDGEOTEPS-----RMGSKSPGNTSQPPAFFSKLTSNSAMVK 2284
Db 2325 SE--YDGESESTEPVHDVETSDTEPTSDAKLKPTSPATPSPESPATEASIVETAPEL 2382
QY 2285 SK-----KQINKKLANHNRNPEYNIQSP--GTEIFNMPAITGTGLMYRSQAOVE 2334
Db 2383 EKEVPEKATEQPELEKETPEKATEQPELEKETPEKATE--QPEL-----EKETPE 2430
QY 2335 HASTNMLEAIIRKALMGKVDQWEESPPLSANAFNPLNASASIPAAAMPITTAAGRSHTL 2394
Db 2431 KATEQPELEKEVT-----DKATEQPEVDEKTTTP-----EPVVKPSLDSTEDESV 2477
QY 2395 TSPGGGG---KAKVSGRPSRRKAKSP-APGLASGDRPPRSVSVHSGD-----CNRR 2442
Db 2478 ESEESADKKDKNKETEEDTDKKEEPEVPVAVVEIQPESEEAIVTGHPLFLPHLASST 2537
QY 2443 TP--LTVRWED-----RPSSAGSTPPFPNPLIMELQAGVNAS-----PPPPGLPA 2486
Db 2538 TTPAVDDRGEEDDEENTVKLSSTSTITESPVSAPTSTTTTASQQQQTITPPYV--- 2594
QY 2487 GSGPLAGPHHA-----WDEE 2501

Db 2595 -----HAPEYEDYDEE 2606

RESULT 54
ABG17148
ID ABG17148 standard; protein; 2759 AA.
XX
AC ABG17148;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17139.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73;
DR N-PSDB; AAS81335.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
FS Claim 20; SEQ ID NO 47507; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2759 AA;

Query Match 3.3%; Score 441; DB 4; Length 2759;
Best Local Similarity 18.4%; Pred. No. 2.1e-15;
Matches 435; Conservative 300; Mismatches 920; Indels 704; Gaps 104;
403 DOQRIKEFNWGLMADPMKYVQKQVNMWMSQEKETFREKFMQHPKXNFGLIASFLEKRT 462
95 DQSAIKNI-----ESAKGRDWDNSPLQWADGTVAETGYHVVFALACS---RKD 141

463 VAECLVYYLTKNENYKSLVRRSYRRRGKSGQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 522
142 ADDTSIYMFYQKGVN-----SIDSWKNA 165
523 KEKEKEAEKEBEKEVEKEDLLKEKTDGDDNDEKAEVASKRKT--ANSQGRKK 579
166 GRVFKDSDFDA-----NDPILKDOTQEWSG-----SATFTSDGKIRLFTYDSGKH 213
580 GR--ITRSMANEANSEEAITPQSAELASM-----ELNESSRWTEEMETAKK 625
214 GKQSLTTAQNVSDDTLKINGVEDHKTIFDGGKTYQNVQOFIDEGNYTSGDNHLLRD 273
626 G--LLEHGRNWSAIARMVG-SKTVSQCKNFYNYKQRNLDEILQOHKLKMEKER--NAR 680
274 PHYVEDKGHKLVFPAQHQGEQKTDQSGESLFNKAYYGGTNFFQSSKLKVKQVPEWNSV 333
681 RKKKKAPAAAEAAFPVVEDEEMEAAGVSGNEEMVEEAALHASGNEVPRGCSGPA 740
334 RVTFNIPREAAERLILAQSNQQQLDLGLISVQIE-GEQAINLALAQNRSDQVRMGPM 392
741 TVNNSDTEIPSPHTEAAKDTGQNGPK-----PP-----ATLGADGPPP--GPPTPPRRT 789
393 GAGNSVRMEA-GFPWASGFIIRMNPNPATVMPGNGVSSSMWAPGPNPELQPTERPAS 451
790 SRAPTEP-----TPASEATCAPTPPAPSP-----SAPPVVPVKEKEEETAAAPV 837
452 QSDAMDPLLSGLHIQQQSHPSGLAPPHHPMQFVSVNRQMNPNPFOLOQQQQQQQQQQ 511
838 EGEBEQKPPAAEELAVDTCKAEPEVKSECTEAEBSGPAKDAEAAEAETGALKAEKE 897
512 QQQ 571
898 GGSGRATTA---KSSGAPQDSDSSATCSADEVDEAEAGGDKNRLSLPRPSELLTPTGDP 954
572 GSLGTWTANQWKKAPLP-----GPMQQLQARPSLAT-----604
955 ASFPQLDLKQLKQRAAIPPTQVTKVHEPPREDAAPTKAPAPAPPPQNLQPSDAPQ 1014
605 -----VQT-----PSHPPPPPPFGSQQAQAHYN 628
1015 -PGSPRGKSRPAPPADKEAFAAQAOKLPGPPCWTGSLPPPPVPPPREVIKASHPADPS 1073
629 FQOMNPGQFTAP-----QWKSLOGP-----SRVPTPLQOPHLTNKSP-ASSFS 672
1074 AFSY-APPGHPLPLGLHDTARVLPRLPPTISNP-----PPLISSAKHSVLERQIGALS 1126
673 SFQSGSPASSPT---VNOQQQMGPRPQ--NNPLPQGFQFPVSSPGRNPMV---QQGNVP 725
1127 QGMSVOLHVPYSEHAKAPVGVMTMGLPLPMDPKLAP-FSGVKQEQLSPRGQAGPESLG 1185
726 PNFWMVQQQPPNQ-----GQSLHPLGLGMPKRLPPGFSA-----GQNFPMQ 770
1186 -VPTAQEASVLRGTALGVSFGGSIITKGIPTSRVPSDAITYRGSITHTGTPADVLYKGTIT 1244
771 QVP-----STATTTPGNS---GAP--QLQANQNVQHAGGQAGAPQNM-----809
1245 RIIGEDSPSLDRGSDSLPKGHVIVYEGKHHVLSYEGGMSVTCQCKEDGRSSSGPPHET 1304
810 -----QVSHGPPNMVQPSLMGIGHNNNNQAGTSGVPQVNLNMQGQPOQGP---857
1305 AAPKRYDMMGEVGRVAISSASIEGLMGRAIPPERHSPHLLKEQHHRIGSITQIGIPRSYV 1364
858 -----SQLMGHQQLV-----SQGMVQO-----QGLINPQNPMILS 890
1365 EAQEDVLRREAKLL---KREGTPPPP-----PPSRDLTE-----AYTKQALGP-----LK 1406
891 RAQ---LMPQGMVNVPPQNGLGPSQRTPPKQMLSSQGPQWMAPHNQMMGQGVLLQ 947
1407 LKPAHGLVATVKEAGRSIHETPREE--LRHTPELPIAPPLKEGSIQTGTPLKYDTGAS 1464
948 QNPMIEQIMTQNMQGNKQNTQONQSNVMPGPAQIMRGFTPNMQGNVQPTGOMSGQMLP 1007
1465 TTGSKKHVDVRLSIGSPRTFPVPHPLDVMADARALERACVEESLKRPGTASSSGGSIA 1524

Db 1008 QCGPVNNSPQVNGIQGVLRPPGSPHMAQQH-----GDPATTANNVDSLSQ 1055
QY 1525 GAPVI-----VPELCKPRQ-----SPLTYEDHG-----APFAG-----HLPR 1556
Db 1056 MFEDVSLQQTVMVPPHVOAQNSASGHSFGHGMSEFNAFSGAPNGNQMSCQNGPFPV 1115
QY 1557 GSPVTWREP--TPRLQEGSLSSKASQDRKLTST-----PREIAKSPHS--TVPEHH 1604
Db 1116 NKDVTLTSPLLVNLQSDISAGHFGVNNKQNTNANKPKKKPKPKKNSQQDLNTPDTR 1175
QY 1605 PHFISPYEH--LLRGVSGVDLYRSHIPLADPTSIPIGI-----PLDAAAAYLP-----RHLA 1656
Db 1176 PAGLEEARQPLPCEQGINLDNSG--PKLPFSNRPPGYPSPQVPEORLPQOMPPQLMQHVA 1234
QY 1657 PNTYPHLYP--PYLIR-----GYPTAALNRQTIINDYITSQQMHNTATATAQRAQ 1708
Db 1235 PPQPQQQOPQLPQQQQPPPPSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 1294
QY 1709 MLRG--LSPRESSIALNYAAGRGIDLSQVPHLPVLVPTTP-----GTPATAMDR 1758
Db 1295 PPRGPNLPSORMPMQSGSVPMVSLQ-----PASVPPSPDKQRMMPVNTPLGNSRK 1350
QY 1759 AYLPTAQ--PFSS-----RHSSPSLPGGPHLTKPTTSSSERERDRDRDR 1806
Db 1351 MVYQSPQNPSSPLAEMASLPEASGSEAPSVPGPNMPSHV-----1393
QY 1807 DRERKSILSTTTVEHAPIRWRCQTEQSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSS 1865
Db 1394 -----VLPQNQLMTGP--KGPSPLSATQATPOOPVNSLPSHGH-----HFNVA 1441
QY 1866 RTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST--SPVRPAATFPFATHCP 1924
Db 1442 PTQTS--REKTPNRSAPRYVYQTPNRPSTPESEISLSPER-----LNAS 1486
QY 1925 LDGVYPTLMPEVLLPKAPVARPERPRADTGHAFIAKPPARSGLEBPASSPSKGSBRPL 1984
Db 1487 IAGLFP-----POINIPLPRLNLRGF-----DOQGLNPTTLKATGOAPS 1529
QY 1985 VPPVSGHATARTPAKNLAHPHASPDPAPPASADPHREKTSQKPFISQLELSRLSYH 2044
Db 1530 TMPSNPAI-----PQTHKLSVVVN-----S 1551
QY 2045 GSSYSPEGVPSVPSPSLTHDKGLPKHLELDKSHLEGELRPKOPGPVKLGEEAHL 2104
Db 1552 GKQNSGATKASPSNS-----RRSPGSSR-----1577
QY 2105 HLRPLPESQSSPLLOTAPGVGHQVVTQAHISEVITQDYTRHHHPQLSAPLPAPLY 2164
Db 1578 KTTSPGRQNSKAPK-----TLASQTNAAALQ-----1605
QY 2165 SFGCASCPLDLRRPPSDLYLPPDHGAPAGS--PHSEGGKRSPENKTSVLGGG---ED 2220
Db 1606 -----NVELPRNLVSPPLANPPVPFGFPNNG--LNPQNTSVSVAAGVGVED 1653
QY 2221 GIBPVGPP-----EGMTPEGHSRAVYP--LLYRDGEQTEPSRMSGKS 2261
Db 1654 NKESLNVPOSDCQNSQRKEQVNIELKAPAEVKAVVEDQSKGQSPDNKLP--1711
QY 2262 PGNTSQPAPFSLKTESNAMSVKKQEIKNKLN--THNRNEPEYNIQPG--TEI--FNMP 2317
Db 1712 -----VEENKNLVSPAMREAPTSLSQLLDNSGAPNVTIKPPGLTDLEVTTP 1757
QY 2318 AITGTGLM-----TYRSQVQEHASTNMGL-----EAILKALMGKYDQWES-----2360
Db 1758 VWSGEDLKASVITPDLQDLSSEKSPNSLNLPHSNLCSLVLHPPELSESSNVAIPPV 1817
QY 2361 -----PPLSANAF--NPLNASASLPAAMPITAAADGRSDHTLTSPGGGKA 2403
Db 1818 MSRPVSSSISTPLPQNQITVFTSNPITTSANTSAALPHLOSALMSTVTPNAGSKV 1877
QY 2404 KVS-----GRP-----SSRKAKSPAPCLA--SGDRPPSVSSV--2433

Db 1878 MVSEGOAQAQSNARPOFITPVPFINSSIIQVMKGSQSPSTIPAAPLTTNSGLMPSPSVAVVG 1937
QY 2434 --HSEGDGDN--RRTPLTNRVMEWRPSSAGSTPPFY-----NPLIMRLQAGVNASPPPPGLPA 2486
Db 1938 PLHIPQNIKSFSSAPVP-----PNAISSSPAENIQTRPLVLSRATPVQLPSP---PC 1987
QY 2487 GSGPLAGPHHAWDEEPKPL 2505
Db 1988 TSSPVV--PSHPVPVQVKEL 2005
RESULT 55
ABB58514
ID ABB58514 standard; protein; 2897 AA.
XX ABB58514;
XX
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 2334.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX OS
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers BW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL02617.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
XX Disclosure; SEQ ID NO 2334; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 2897 AA;
SQ
Query Match 3.3%; Score 440; DB 4; Length 2897;
Best Local Similarity 18.8%; Pred. No. 2.5e-15;
Matches 526; Conservative 316; Mismatches 1054; Indels 904; Gaps 119;
QY 60 QPQRRRPSLLS-----EFQPNERSQELHLRPESHSYLPELGKSEMEF 102
Db 326 QQKKHPTLLDTKRIKKPSRYESTKDEIBESGEDEEE---BDEEDSELDE--GEE-ED 380
QY 103 IESKRPRLELLPD-----PLLRPSPLATGQAGSEDLTKDRSLTKGL-----EP 147
Db 381 VDSAAADIEDDEEDTGTGPEDPT--IVVQGGSGLDCEAIPYNLGNFDYEDDLEIGA 439

Db 2341 LVPSSADLAR-----FYQVQANQQPI-PAVPGSR-----PSSTSGPRH----- 2379
QY 1923 GTLDGVYPTLMEVLLPKAPRVARPERPRADTGHAFLAKPPAPSGLEPASSPSKSGSEPR 1982
Db 2380 -----LLRQMPGGLP-----PHASLRPTYGGPPPLRGSGPTSTSTTSNR 2423
QY 1983 PLVPPVSGHATARTPAKNAPHASGPDPP-----APPASADPHREKTSKPF81 2033
Db 2424 PAY--LHG-----AEHGGSPGPMGGVSSGPPPARHATPHLNPYRAPPI-- 2467
QY 2034 QELESLGVLGHSYSPE-GVEPVSPVSPSLTHDKLPHLELDKSHLEGLRKPQG 2092
Db 2468 -----YGNPNYSPRLGGAPGT-----GSMR-----PG 2489
QY 2093 PVK-LGEAAHLP-----HLRPLPSQPS-----SSPLLOTAPGVKGHQRVVTLAOHISEVIT 2144
Db 2490 ADVYAGPRGYSYGYPPPPPLSTSAHATSVIVSAP-----HTLTPTNHSVPTLT 2543
QY 2145 QDVT--RHHFQQLSAPLAPLYSPGASCPVL-----DLRRPPSLDLYLPPD-----HGAPA 2194
Db 2544 HGKTPPQOTPTQSSGPPA-----AAPPPTITSETSSHKPPLASVITSKLTLEAYPI 2597
QY 2195 RGSFHS-----EGKRSPEPNKTSVLGGEDGIEPVSPGEMTEPGHRSADVPLLY 2246
Db 2598 RKSPIAVADVSGPAEPTRSAPABEDSGSAHDTRAPSSA-----TGTAVVGEFS 2648
QY 2247 RDGEQTEPSRMGSKSPGNTSQPPAFTSKLTESAMVSKKQBIKKNLTHNNEPEYNI 2306
Db 2649 GSGNANQAHCTGPTPRELQ-----SKL-ELEQOSKLEREQEPSKL-----EEL 2694
QY 2307 SQFGTEFINMPTITGTGLMYRSQAVQERHASTMNGLEAII 2346
Db 2695 EQOSKLERAQREGRGREWNTRELAQEQSKLELQEQSI 2734

RESULT 56
AAE13839
ID AAE13839 standard; protein; 4019 AA.
XX AC AAE13839;
XX DT 26-FEB-2002 (first entry)
XX DE Human lung tumour-specific protein SCC2-29.
XX DE Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antinease-therapy; vaccine; immune response; lung cancer; SCC2-29.
XX OS Homo sapiens.
XX PN WO200172295-A2.
XX PD 04-OCT-2001.
XX PF 28-MAR-2001; 2001WO-US009991.
XX PR 29-MAR-2000; 2000US-00538037.
XX PR 05-JUN-2000; 2000US-0058937.
XX PR 18-AUG-2000; 2000US-00640878.
XX PR 22-SEP-2000; 2000US-0234517P.
XX PR 01-NOV-2000; 2000US-00704512.
XX PR 14-DEC-2000; 2000US-00738973.
XX PR (CORI-) CORIXA CORP.
XX PA Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
PI WPI; 2001-639201/73.
DR N-PSDB; AAD23451.
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer.

XX PS
XX CC The invention relates to isolated lung tumour-specific proteins and their
CC corresponding cDNA molecules. Lung tumour-specific proteins and their
CC antigen-presenting cells are useful for stimulating and/or expanding T
CC cells specific for a tumour protein, and for inhibiting the development
CC of cancer. The invention also relates to a composition useful for
CC stimulating an immune response, and for treating cancer. The lung tumour
CC specific oligonucleotide is useful in gene therapy and for diagnosis,
CC detection and treatment of lung cancer. The present sequence is human
XX lung tumour-specific protein
SQ Sequence 4019 AA;
Query Match 3.3%; Score 436; DB 4; Length 4019;
Best Local Similarity 18.3%; Pred. No. 6.5e-15;
Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
QY 34 HTDVGLELYQHHSRDYASHLSPGSIIP-----QRRRPSLLSEFPQGNERSQELHLPESH 90
Db 513 HSDIGPVT-----DFSSLPQPNVQSSRP--LSEQLDGLSPELDGMVTDGA 559
QY 91 YL-----PELKSMEFIESKRPLELLPDLPLRPSPLLATGQAGSBDLTKDRSLTGK 144
Db 560 ILGLKYKIPELGKGVDEL-----FTAVLSPANT----- 588
QY 145 LEPVSPSPPHDTPLELVPPLSKKEELIQNDVRVDREITWVEQOIKLKKKQOQLEEEA 204
Db 589 -QFTPLPQP-----PPTQLLP-----IHNQDAFSR-MPLMNGLI-----G 632
QY 205 AKPPEPEKPVSPPTIESKHSRLVLIIDENRKAEEAHRILEGLGQVVELPLYNQPSDTR 264
Db 623 SSPHLPHNSLPPGSLGTESAQAQSSYPDARDKNSAFN-----PMASDNNNS- 669
QY 265 QYHENIKINQAMRKLLILYFKRRNHARKOWKQFCORYDQLMLEALEKKYVERIENNRRRA 324
Db 670 -----WTSS-----APTVEG-ENDTWSNA 687
QY 325 KESKVEYVEKQFPPEIRKQRELQERMQSRVQSGLSMSAARSEHEVSEIIDGLSEQEN 384
Db 688 QRSTLK--WEK----- 696
QY 385 LEKOMRQLAVIPPLMLYDADQRIKFINMGLMADPMVKYKDRQVMNMWSEQKETREKF 444
Db 697 -EALCEMATVAPLY-----TWINPENKEEPPD--WTRVQIAKLWK----- 739
QY 445 MOHPNFGLIASFLERKTVAECVLYYYLTKNENYKSLVRSYRRRGKSGQOQOQOQOQOQ 504
Db 740 -----ASSQERAP-----YVQKARDN-----RAALRINKVQMSNDSMKRQ 775
QY 505 QOQOQOQPMRPSQOEKDEKE 564
Db 776 QQSDIDPSRIDSELSFKDPLKQESHEQE----- 805
QY 565 ASKGRKTANSQGRKGR--TRSMANEANSEBAITPQSAELASMEELNESSRTEEMET 622
Db 806 -WKFRQMRQKSKQAKIETATKLEQVKNQEQ-----QOQOQOQSGSHL-----LVQSGSDT 856
QY 623 AKKGLLEHGRNWSAIRMVSKTVSQCKNFYNYKQRNLDLILQOHLKMEKERNARRK 682
Db 857 PSSGI-----QSPLTPQPGNGMNSPAQSF-----HKLFTKQ----- 888
QY 683 KKKAPAAASEEAAF-----PPV-----VEDEMEASGVSGNEEMVEAEALHASGN 729
Db 889 PFSTPTSTSDSDVFKVQAPPPPPPPAPSRIPIDSLSQ----- 936
QY 730 EYPRGCSGPATVNNSSDTEIPSPHTEAAKQDQNGKPPATLGDGPPGPPPTPP--- 786
Db 927 -----QTSQPPSPQVFS-----PGSSNSRPPSPMDPYAKMVGTPRPPVPG 966
QY 787 -----RTSRAPISP-TPASEA-----TGAPTPPPAPP-----SPSAPPPVVPK 824

Disclosure; Page 309-318; 378pp; English.

Db 967 HSFRRNSAAPVENCCTPLSVSRPLQMNETTANRPSFVRDLCSSTNTNDPVAKPDP- 1025
QY 825 EKEEETAAPVVEEGEOKPAAAEELAVDTGKAEPVKSECTEEAEPAKGDAAEAE 884
Db 1026 -----RPVMTDFPKSLGLSRSPVISEQT-----AKG---PIAA 1056
QY 885 ATAEGALKAEBGGGRATTAKSGAPODSSSATCSADEVDEAGGDKNLLSPRSL 944
Db 1057 GTSDHFTKE-----SPRADVFQORIP--DSYARPLLTPAPLDSGPGFKTP--MQPPSS 1108
QY 945 LPTGDPANASPOKPLDLKOLKQRAAAIIPPTQVTKHEPPREDAAATKAPAPPPQN 1004
Db 1109 QDPYGSV--SOAGRLSVD---PYERPALTPRIDNFNSHNQNDPYSQPLTPHFAVNES 1163
QY 1005 LQPEDAPOQPGSPRGSKR-----SPAPPADKEAFAEAQKLPGLPPCWTSLPFP 1056
Db 1164 FAHPSRAFSQGTIISPTSQDPYQPPCTPRFVVDYSYSSGSTARNTDYPQPGTTPR 1223
QY 1057 --VPREVITAKSHAPDPSAFS-----YA--PPGHPLP--LGLHDTARPVLP 1097
Db 1224 TTVDP--YSQOPOTPRSTQTDLFVTPVTNQRHSDPYAHPGTGPRGISVPYQPPATP 1280
QY 1098 RP-----PTI--SNPPPLISSAKH-----PSVLERQIGAISQMSVOLHVYS 1138
Db 1281 RPRISEGFTRSSMTRVLMNPQDFFLOAAQNRGAPALFGLVLRPPDTCQ----- 1329
QY 1139 EHAKAQVGPVTWGL-----PLPMDPKKLAPSGVKQ-----EOLSPR 1175
Db 1330 --TPRPPGP--GLSDTFSRVSPAARDPYDQSPMTFRSQSDSFGTSQTAHDVADQPRG 1384
QY 1176 GQAGPPESLGVPTAQBASVLRGALGSVFGSITKGIPT----- 1215
Db 1385 SEGSCASSNPMHSQOQFSG--VSQLPFGVPTSGVTDTQNTVNAQAADTEKLROKQL 1442
QY 1216 -----RVPSDS--AITYRGSTHGTADV-----LYGHTIRII 1247
Db 1443 REILQOQOQKTIAGRQSGQSPAVPHFGPLQHWQPNVNAQFTRPPPPYFGNIR-- 1499
QY 1248 GEDSPRLDRG--REDSLPLKGH-----VIECKKGH--VLSYEGQMSVTQCKEDGRSS 1297
Db 1500 ---SPVAPLGPRIYVFPKQDQGPYPDVASMGMRPHGFRFGPGSGHGTMPQOE---RF 1553
QY 1298 SGPPHET---AAPKR---TYDMEGRVGRATISSASIEGLMGRATIPRERHSHHLK--EQ 1348
Db 1554 LVPPQOIQSGSVSPQLRRSVSDM-----PRLLNSQMNVPGL---PQHFSQSLPVQO 1605
QY 1349 HHIRGI-----TQIGPRSYVEAQEDYLREAKLLKRGTPPP--PPPS 1390
Db 1606 HNILQOAYIELRHRAPDGRQRLFPFSPAPGSVVEASSNL--RHGNFTPRDPFGPRHTDPM 1663
QY 1391 RDLTEAYKTQALGPLKLPKPAHEGLVATVKEAGRSIH--BIPRELRHT--PELPLAPRPL 1446
Db 1664 RPPQGLPNQ-----LPVHPDLEQVPSQEQGHSVHSSWMWRTLNHPILGGESEAPLST 1719
QY 1447 KEGSITQGTPLKYDTCASTGSKKHVRSILGSPGRTFFPVHPLDMADARALE--RACYE 1505
Db 1720 SVFSETSDNLQITTPQSDGLEKLDSD-----PSVKELDQ--KOLEGVVVKDLD 1769
QY 1506 ESLKSPGTASSSGGSJAR-----CAPVIV-----PELCKPQPSPL 1541
Db 1770 EDLENJ--NLDTEDGKVVELDLDNLETNDPNLDDLLRSGEFDIAYTDPFLDMGDKKSM 1827
QY 1542 TVEDHGAPFAGHLPRGSPVMTREPTPRLOEGS--LSSSKASQDRKLTST----- 1588
Db 1828 FNEELDLPIDDKL--DNQCVSVPEKKQEENKTLVLSDKHSQKSTVNEVKEVLSN 1885
QY 1589 -----PREIAKSHSTVPEHHP-----HPISP--YEHLLR----- 1616
Db 1886 SKVESKETENDKNDVTPCSQASASDLNDGDKTSLHPCDPLFEKRTNRETAGPS 1945
QY 1617 -----GVSVDLYRSHIPLA---FQPTSI--PRGIPLDAAAAYLPR 1653
Db 1946 ANVIOASTQLPAQDVINSGCITGSTPVLSSL--LANEKSNDSDIRPSGSP----- 1993

QY 1654 HLAHPNTYP-----HL--YPPYLIRGYPDTALENRQTIINDYIT--SQOMHHTATAMAQ 1705
Db 1994 ---PPPTLPASPSNVSSLPPIA---PPGRVLDN---AMNSNVTVVSRVNH---VFSQ 2040
QY 1706 RADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPPPTGTPATAMDRLAYLPTAP 1765
Db 2041 GVVNPNGLIPQOST--VNHSLG-----TGKPAT-----QTGP 2070
QY 1766 QPFSRSHSSPLSPGPGTHLTKPTTTSSEER-----DRDRERDRER 1810
Db 2071 Q--TSQSGTSSMS--GPQOLMIPQTLAQONRERPLLEEQPLLLQDLQERQEQOQO 2126
QY 1811 EKSIIITSTT--TVEHAPIWRPGTEQS----- 1834
Db 2127 MQAMIRQORSEPPFNIDFDAITDPIKAKWALKGINKVMAQNNLGMPPMVMRPFPMQ 2186
QY 1835 --SGSGSGSGSGSGSSRASHAHQHSPISP-----RTQDAL 1871
Db 2187 VVTGTQNSGQNLGPQAIPQDGSITHQISRPNPNFGPGFVNDSORKQYEEWLQETQQL 2246
QY 1872 QORPSVL-----HNTGMKGIITAVEPSKPTVLRSTSTSSVVRPAATPPATHCPLGGTL 1925
Db 2247 QMOQKYLEEQIGAHRSKKAL-----SAKQRTAKKAGREFPEDEAQLKHVTE----- 2294
QY 1926 DGVIYPTLMEPVLLPKXAPRVARPERPRADTGHAFIAK-----PPA-----RSGLE 1970
Db 2295 -----QOSMVQKQLEQIRKQKQHEAELIEDYIRKQOQOCAMAPPTMMPSVQOPPLI 2346
QY 1971 PASSPSKSEPR--PLVPP-----VSGHATTARTPA-----KNLAPHASDPDP-- 2012
Db 2347 PGATPTWTSQPTFPMVPOOLOHQOHTTVISGTSVPRMPSLPGWOPNSAPAHLPNPPRI 2406
QY 2013 -----APPASADPHREKTSQKPS----- 2032
Db 2407 QPIIAQLPIKCTPAGTVSNANPQSGPPRVFEDDNNPFSFQERERKEBLREQERQ 2466
QY 2033 -----IQELELSRGLYHGSYSYSP-----EGVEVPSPVS 2061
Db 2467 RIQLMQEVDQRALQORMEQMGVMEGSEISSRTSVSQIPFYSSDLPCDFMQPLGLQ 2526
QY 2062 PSLTHDKGLPKHLEL-----DKSHLEGLRP-----KQPGVKLGGEAAHL 2104
Db 2527 S-----PQOQOQGVLOQOQNIQOQSINSFSTQTFMOTNERRQVGPSPVPSPIP 2578
QY 2105 HLRPLPES-----OPSSSPLLQTAGVKGHQVWVTLAQHISEVITQ 2146
Db 2579 VGSNPFSSVKQGHNLGSGTSFOQSPVRFSFTPALPAAPV-----ANSSLPCCQD 2628
QY 2147 YTRHHPQQLSAPLAPLSPFGASCPCVLDLRPPDLPLPPDHGAPARGSPHSEGGKRS 2206
Db 2629 STITHG-----HSYPGSTQSLIQLY---SDII-----PEEKKKKR 2661
QY 2207 PEPNKTVLGGEDGIEBVPSPPEGMT---EPGHSRAVYPLLYRDOEQTEPERMSKSPG 2263
Db 2662 TRKKKRD--DDAESTKAPSTPHSDITAPPTPGISETTSTPAVSTPSLPPQADQSVFV 2719
QY 2264 NTSQPPAPFSKL--TESNSAMVKSQKQINKNLNTHRNEPEYNIQPGTEIFNMPAITCT 2322
Db 2720 GPSTNMAAGQJCTELENKLP-----NSDFSQATPNQOITYANSEVDKLSMETPAKT-- 2770
QY 2323 GLMTRYSQAVQASHASTNMGLEAIIRKALMGKYDQWEEPSPLSANAFNPLNASLPAAMP 2382
Db 2771 -----EEIKLEKAETESCPG-----QEEPKLEQNGSKVEGNA---VACP 2807
QY 2383 IYAADGRSDHUTLSP---GGGKAKVSGRPSRKAASPAAGLASGDRPPSVSVSEHSGDC 2439
Db 2808 VSSAOS--PPHSAGAPAAKGDGSGNELLKHLKNKKSSS-----LLNQKPE--GSI 2859
QY 2440 NRTPLTNRVWEDRPSAGSTPPFNPLIMRLQAGVMASPPPPGLPAGS----- 2488
Db 2860 TKONKL---VEKONFAEGLQT-----LQAQMGGFGCGNQLPKTDGSGSETKKQSRKTQ 2910

QY 2489 --CPLAGPH--HAWDEEPKPLCSQYETLS 2514
 Db 2911 RTGEKAAPRSKGRKDEEKKQAMYSSTDTFT 2941

RESULT 57
 ADD66733
 ID ADD66733 standard; protein; 4019 AA.
 AC ADD66733;
 XX 15-JAN-2004 (first entry)
 DT Human lung tumour-specific related protein, SEQ ID No 425.
 XX expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;
 KW human; lung tumour-specific.
 XX Homo sapiens.
 OS WO200292001-A2.
 PN 21-NOV-2002.
 PD 10-MAY-2002; 2002WO-US014975.
 XX 11-MAY-2001; 2001US-00854133.
 PR (CORI-) CORIXA CORP.
 XX Lodes MJ, Wang T, Fan L, Algate PA, Mcneill PD;
 PI WPI; 2003-120592/11.
 XX New polynucleotide and polypeptide, useful for preparing a composition
 PT for diagnosing, treating or preventing cancer.
 XX Disclosure; SEQ ID NO 425; 494pp; English.

The invention relates to a novel isolated polynucleotide comprising one
 of 32 47-6080 base pair sequences, given in the specification, or their
 complements or degenerate variants, at least 20 contiguous residues of a
 sequence in, or having at least 75 or 90 % identity with the isolated
 polynucleotide, or that hybridise with the polynucleotide. The invention
 further comprises: an isolated polypeptide; an expression vector
 comprising the polynucleotide operably linked to an expression control
 sequence; a host cell transformed or transfected with the expression
 vector; an isolated antibody or its antigen-binding fragment that
 specifically binds to the polypeptide; a method for detecting the
 presence of a cancer in a patient; a fusion protein comprising the
 polypeptide; an oligonucleotide that hybridises to the isolated
 polynucleotide under moderately stringent conditions; a method for
 stimulating and/or expanding T cells specific for a tumour protein; an
 isolated T cell population; a composition comprising a first component
 consisting of carriers and immunostimulants and a second component; a
 method for stimulating an immune response in a patient; a method for
 treating cancer in a patient; a method for determining cancer in a
 patient; a diagnostic kit comprising at least one oligonucleotide or
 antibody and a detection reagent comprising a reporter group; and a
 method for inhibiting the development of cancer in a patient. The
 compositions of the invention have cytostatic activity and can be used to
 create a vaccine. The isolated polynucleotide is useful for preparing a
 composition for diagnosing, treating or preventing cancer. This sequence
 represents a human lung tumour-specific protein relating to the
 invention.

Sequence 4019 AA;
 Query Match 3.3%; Score 436; DB 7; Length 4019;
 Best Local Similarity 18.3%; Pred. No. 6 5e-15;
 Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
 34 HTDVGLEYQHHSRDYASHLSGSIITQP---QRRRPSSLSEFQGNERSQELHLRPESH 90

Db 513 HSDIGVTD-----DFSSLPQPNVNSSRP--LSEBOLDGILSPELDQMTDGA 559
 QY 91 YL-----PELGKSEMEFTESKRPRLELLPDLPLLRSPPLATCQAPAGSEDLTKDRSLTGK 144
 Db 560 ILGKLYKIPELGGKVEDL-----FTAVLSPANT----- 588
 QY 145 LEPVSPSPPTDPELVLPPRLSKKEELTONMDRVDRITMVEQQISKUKKQQQLEBEA 204
 Db 589 -QPTPLPQP--PPTQLLP-----IHNQDAFSR-MPLMNGLI-----G 622
 QY 205 AKPPEPEKVPSPPIESKHSLSVQIIYDENRKKAAEAHRLLEGLGQVVELPLVNQPSDTR 264
 Db 623 SSPHLPHNSLPPGSLGTFSATQSSYPDARDKNSAFN-----PMASDNNNS-- 669
 QY 265 QYHENIKINQAMRKLLILYFKERNHARKQWKQFCQRYDQLMEALEKKYVERIENNRRA 324
 Db 670 -----WTSS-----APTVEG-ENDTMSNA 687
 QY 325 KESKREYVEKOPPEIRKQRELOERMQSVRGQSGLSMSAARSEHEVSEIIDGLSEQEN 384
 Db 688 QRSTLK--WEK----- 696
 QY 385 LEQMRQLAVIPMLYDADQRIKFINNGLMADPMKYVKDRQVMNMSEQEKETREKF 444
 Db 697 -EALGEMATVAPVLY---TNINPENLKEEFPD--WTRVKQIAKLRK----- 739
 QY 445 MQHPKNFGLIASFLERKTVAECVLYLYLTCKENYKSLVRRSYRRRGKSGQQQQQQQQQQ 504
 Db 740 -----ASQERAP-----YVOKARDN-----RAALRINKVQMSNDSMKRQQ 775
 QY 505 QQQQQQMPRSSQBEKE 564
 Db 776 QQSDIDPSSRIDSELPKDLKQRESEHEQE----- 805
 QY 565 ASKGRKTANSQRRKGR1--TRSMANEANSEERAITPQSAELASMLNBSRRWTESEMET 622
 Db 806 -WKFQRMQKSKQAKIEATQKLEQVKNQEQ--QQQQQQQFGSQHL-----LVQSGSDT 856
 QY 623 AKKGLEHGRNNSATARMVGSKTYSCKNFYFNYKKRQNLDELITQHKLMKERNARRK 682
 Db 857 PSGI-----QSPLTPQNGNMSPAQSF-----HKELFTKQ----- 888
 QY 683 KKKAPAAASEEAAAF-----PPV-----VEDEMEASGVSGNEEMVEEAEALHASGN 729
 Db 889 PESTPTSTSSDDVFVKPQAPPPPPAPSRIPIDSLQA----- 926
 QY 730 EVPRGECGPATVNNSSDTEIPSPHTEAAKDTGQNGPKPATLGADGPPPGPTTP-- 786
 Db 927 -----QTSQPPSPQVFS---PGSSNSRPPSPMPDYAKMVGTPRPPVPG 966
 QY 787 ---RRTSRAPTEP-TPASEA-----TGAPTPPPAPP-----SPSAPPVVPVK 824
 Db 967 HFSRRNSAAPVENCPTLSVSRPLQMNETTANRPSFVDRDLCSSTTNDNDPYAKPDTP- 1025
 QY 825 EKEEETAAPVVEGEQKPPAAEALAVDTGKAEPPVKSECTEEAEAGPKAGKDAEAE 884
 Db 1026 -----RPVMTDQFKSLGLSRSPVVSQEI-----AKG---PIAA 1056
 QY 885 ATAGGALKAEKGGGGRATTAKSGAPQDSSATCSADEVDEAGGDKNLLSRPSSL 944
 Db 1057 GTSDHFTK-----SPRADVFQRIIP--DSYARPLLTAPLDSGFGPFTKTP-MQPPSS 1108
 QY 945 LPTGDPANASPOKELDLKQLKQRAAAIPPIQVTKVHEPPRPPDAAPTAPAPPPPPON 1004
 Db 1109 QDPYGSV-SQASRRLSVD-----PYERPALTTPRIDNFSHNQSNNDPYSQPLTPHPAVNES 1163
 QY 1005 LOPESDAPQPGSSPRGKSR-----SPAPPADKEAFAAEAQKLPDGPWCWTSGLPFP 1056
 Db 1164 FAHPSEAFSQPGTISRPTSDQPYSPQPGTFRPVVDSYSQSSGTARSTNTDPSYSPGCTPRP 1223
 QY 1057 --VPREVITKASHPADPSAFS-----YA-PPGHPLP-LGLHDTARPVL 1097

Db 1224 TTVDP---YSQQPQTPRPSTQTDLFVTPVTNQRHSDPYAHPECTPRPGISVFPSPAPT 1280
Qy 1098 RP-----PTI-SNPPLISSAKH-----PSVLERQIGAISQMSVOLHVPYS 1138
Db 1281 RPRISGFTSSWTRVPLMNPQDFLOAQRNGPALPGLVPRPDTCSQ----- 1329
Qy 1139 EHAQAVGPVTMGL-----PLPMDPKKLAPFGVKQ-----EQLSPR 1175
Db 1330 --TPRPPGP--GLSTDFSRVPSAARDPYDQSPMTPRSQSDSFGTSQTAHDVADQPRFG 1384
Qy 1176 GQAGPESLGVPYTAQASVLRGALGSPGSGITTKGIPST----- 1215
Db 1385 SEGSCASSNSPMHSQCGQFSG--VSQLEGPVPTSGVTDQTNTVMAQADTEKLQRQKL 1442
Qy 1216 -----RVPSDS-AITYRGSITHTGTRADV-----LYKGTITRII 1247
Db 1443 REILIOOQOKKIAGROEGSQSPAVPHGPLOHQWOPENVNOAFTRRPPPPYGNIR---- 1499
Qy 1248 GEDSPRLDRG-REDSLPKGH-----VIYEKGKGH--VLSYEGGMSVTQCKEDGRSS 1297
Db 1500 ---SPVAPPLGPYAVFPKQRCGPYPDVASMGMRPHGFRFGPGGSHGTMPQOE---RF 1553
Qy 1298 SGPPHET---AAPKR---TYDMEGRVGRRAISSAIEGLMGRATPERRHSHHLK-EQ 1348
Db 1554 LVPPQIQGSGVSPQLRRSVSDM-----PRLLNSQMNPNVGL---POHFSQSLPVQ 1605
Qy 1349 HHIRGSI-----TQGI PRSYVEAQEDYLREAKLLKRGSTPPP---PPPS 1390
Db 1606 HNILGOAYIELRHAPDGRQRLPFSAPPQSVBASSNL--RHGNFIPRPDFFGFRUTDPM 1663
Qy 1391 RDLTEAYKTQALGLPKAHEGLVATVKEAGRSIH--BIPRELRHT--PELPLAPRPL 1446
Db 1664 RRPQGLPQO---LPVHPDLEQVPPSQEQGHSHSSVWVETLHNLPGGEFSEAPLST 1719
Qy 1447 KEGSITQGTPLKYDTGASTGSKKHVRSIGSPGRTFPPVHPDVMADARALE-RACYE 1505
Db 1720 SVPSSETSDNLQITTPQDGLBEKLSDD-----PSVKELDV-KOLEGVEVVDLDD 1769
Qy 1506 ESLKSRPGTASSGGSIAR-----GAPVIV---PELGKPRQSPL 1541
Db 1770 EDLENL--NLDTEGKVVELDTLDNLETNDPLDLDLRSGEFDIIAYTPELMDGDKKSM 1827
Qy 1542 TYEDHGAFFAGHLPRGSPVPMREBTPRLQEGS--LSSSKASQDKLST----- 1588
Db 1828 FNEELDLPIDDKL--DNQCVSVPEPKKEQENKTLVLSDKHSPOKSTVTNEVKTEVLSPN 1885
Qy 1589 -----PREIAKSPHSTVPEHIP-----HPISP--YEHLR----- 1616
Db 1886 SKVESKETEKNENKDNVDTPCSQASAHSDLDNGEKTSILHPCDPLFEKRTNRETAGPS 1945
Qy 1617 -----GVSGVDLYRSHIPLA--FDPTSI-PRGILPLDAAAAYVLP 1653
Db 1946 ANVIQASTQLPAQDVINSCGITGSTVLSLSL-JANEKSDNSDIRPSGSP----- 1993
Qy 1654 HLAFNPTYP-----HL--YPPYLIRGYPDPAALENRQTIINDYIT-SQQMHNTATAMAQ 1705
Db 1994 ---PPPTLPASPSNHVSSLPFFIA---PPGRVLDN---AMNSVTVVSVNH---VFSQ 2040
Qy 1706 RADMLRGLSPRESLAINAAGPRGIDLSQVPHLPVLVPPPTGCTATAMDRLAYLPTAP 1765
Db 2041 GVQVNPGLIPQGST--VNHSLG-----TKGPAT-----QTGP 2070
Qy 1766 QPFSSRHSSSPLSGPGTHLTKTPTTSSRRER-----DRDRDRDRDR 1810
Db 2071 Q--TSQSGTSSMS--GQQUMIQTUAAQNRERPLLEBQPLLDLLOEREOQQORQ 2126
Qy 1811 EKSILSTT-----TVEHAPIMRPGTEQS----- 1834
Db 2127 MQAMIRQSEFPFNIDFALITDPIKKAQVVALKINKVMAQNGLMPPMVMSRFFPMGQ 2186
Qy 1835 --SGSSGSSGGGSSSRPASHAHQHSFISP-----RTQDAL 1871
Db 2187 VVTGTQNSEQNLPQAIPODGSITHQISRPNPNFPGFVNDSPQRKQYEWLQETQQL 2246

Qy 1872 QORPSVL-----HNTGMKGIIITAVEPSKPTVLRSTSTSSSPVRPAATFPFATHCPGLGTL 1925
Db 2247 QWQKYLBEOICAHRSKKAL-----SAKQRTAKKAGREFPEEDAQLKHVTE----- 2294
Qy 1926 DGYYPTLMEPVLLPKAARVARPERPRADTGHAFIAK-----PPA-----RSGL 1970
Db 2295 -----QOSMVQKLEQIRKQKEHAELIEDYRIKQOQOCAMAPPTMPSVQOPPLI 2346
Qy 1971 PASSPSKSEPR-PLVPP-----VSGHATARTPA-----KNLAPHASDPDP-- 2012
Db 2347 PGATPTWSQPTFPVMVPOOLOHQHQTWTISGHTSVRMPSLPGWOPNSAPAHPLPLNPRI 2406
Qy 2013 -----APPASADPHREKTQSKPFS----- 2032
Db 2407 QPPIAQLPIKTCTPAGTVSNANPOSPPRPFVEFDNNPFSSFOERERKERLRBQERQ 2466
Qy 2033 -----TOELERLSLGVHGSYSYP-----EGVEPVSPVS 2061
Db 2467 RIQLMQEVDQRALQORMEMEQHGMVGEISSRSTSVSQIPFYSSDLPDCDFMQLGPLQ 2526
Qy 2062 PSLTHDKGLPKHLEEL-----DKSHLEGELRP-----KQPGVKLGGAAHL 2104
Db 2527 S-----POHQOQMGVLOQONIQGSINSFSTQTFMTNRRQVGPSPFVDPSPST 2578
Qy 2105 HLRPLPES-----QFSSSPLLQTPAGVKGHQVRVTLAQHISEVITQD 2146
Db 2579 VGSNPFSSVKQGHGNSLGTSTFQOSPVRPFTPALPAAPV-----ANSSLPCGD 2628
Qy 2147 YTRHHPPQOLSAPLAPLAPLAPLAPLAPLAPLAPLAPLAPLAPLAPLAPLAPLAPLAPL 2206
Db 2629 STITHG-----HSYPGSTQSLIOLY---SDII-----PEEKKKKR 2661
Qy 2207 PEPNKTSLVGGEDGIEPVSPPEGMT---EPGHSRSVAVVPLLYRDEQTEPSPRMGSKSPG 2263
Db 2662 TRKKRDR--DDAESTKAPSTPHSDITAPPTPGISETTSTPAVSTPSELFPQADQESVEPV 2719
Qy 2264 NTSQPPAPFSL--TESNSAMVSKKQEIKNKLNTHNRNEPEYNI SQGTETIFNMPAITGT 2322
Db 2720 GFSTPNMAAGLCTELENKLP-----NSDFSQATPNQOQTYANSEVDKLSMETPAKT-- 2770
Qy 2323 GLMTRYSQAVQEHASHNTMGLEAIIRKALMGKYDQWEESSPPLSANAFNPLNANASLPAAP 2382
Db 2771 -----BEIKLEAETESCFG-----QEEPKEEQNGSKVEGNA---VACP 2807
Qy 2383 ITAADGRSDHILTSP---GGGKAKVSGRPSRKAAPGLASGDRPPSPSVSHSEGDC 2439
Db 2808 VSSAQSG-PPHSAGAPAAKGDGSGNELLKHLKNKKSSS-----LLNQKPE--GSI CSEDDC 2859
Qy 2440 NRTPLTNRVWEDRPSAGSTPPFPYNPLIMRIQAQVMASSPPPPGLPAGS----- 2488
Db 2860 TKDNKL---VEQNPAEGLQT-----LGAQMGGFGCGNQLPKTDGSETKKQSKRTQ 2910
Qy 2489 --GPLAGPH---HAWDEBPPLCSQYETLS 2514
Db 2911 RTGEKAAPRSKKRKKDEEEKQAMYSTDTFT 2941

RESULT 58

ADE87987

ID ADE87987 standard; protein; 4019 AA.

XX ADE87987;

XX AC ADE87987;

XX DT 29-JAN-2004 (first entry)

XX Human lung tumour antigen polypeptide #92.

XX DE Human; lung tumour antigen; cancer; lung cancer; CD8+; T cell;

XX KW Human; lung tumour antigen; cancer; lung cancer; CD8+; T cell;

XX KW immune response; immunostimulant; cytostatic.

XX XX Homo sapiens.

XX OS

PN US2003118599-A1.
 XX 26-JUN-2003.
 XX 10-MAY-2002; 2002US-00144649.
 XX 02-APR-1999; 99US-00285323.
 PR 09-AUG-1999; 99US-00370838.
 PR 30-DEC-1999; 99US-00476235.
 PR 03-MAR-2000; 2000US-00518809.
 PR 29-MAR-2000; 2000US-00538037.
 PR 05-JUN-2000; 2000US-00588937.
 PR 18-AUG-2000; 2000US-00640878.
 PR 20-SEP-2000; 2000US-00667170.
 PR 01-NOV-2000; 2000US-00704512.
 PR 14-DEC-2000; 2000US-00738973.
 PR 11-MAY-2001; 2001US-00854133.
 XX (CORI-) CORIXA CORP.
 PA Algate PA, Lodes MJ, Wang T, Fan L, Mcneill PD;
 PI WPI; 2003-897103/82.
 DR N-PSDB; ADE87984.
 XX
 PT New polynucleotides encode lung tumor antigens and are useful to
 PT stimulate an immune response or detect or treat a cancer in a patient,
 PT particularly lung cancer.
 XX
 PS Disclosure; SEQ ID NO 425; 63pp; English.
 XX
 CC The invention relates to polynucleotides encoding lung tumor antigens.
 CC The invention also relates to the polypeptides encoded by the
 CC polynucleotides, isolated antibodies or antigen-binding fragments that
 CC specifically bind the polypeptides and a method for detecting cancer in a
 CC patient, comprising obtaining a biological sample from the patient,
 CC contacting the sample with a binding agent that binds a polypeptide of
 CC the invention, detecting in the sample an amount of polypeptide that
 CC binds to the binding agent, and comparing the amount of polypeptide to a
 CC predetermined cut-off value. T cells specific for a tumor protein can be
 CC stimulated and/or expanded by contacting the T cells with a polypeptide,
 CC polynucleotide or an antigen-presenting cell that expresses a
 CC polypeptide. Cancer development can be inhibited by incubating CD4+
 CC and/or CD8+ T cells isolated from a patient with a polypeptide,
 CC polynucleotide or an antigen-presenting cell that expresses a
 CC polypeptide, so that the T cells proliferate. The invention is used to
 CC stimulate an immune response or to detect or treat a cancer in a patient,
 CC particularly lung cancer. This sequence represents a human lung tumor
 CC antigen polypeptide of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification but was obtained in
 CC electronic format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 4019 AA;
 Query Match 3.3%; Score 436; DB 7; Length 4019;
 Best Local Similarity 18.3%; Pred. No. 6-5e-15;
 Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
 34 HTDVGLLEVOHHRDYASHLSQSIQTP---QRRPSLLSEFQFNGRQELHLRPESH 90
 513 HSDIGFVTD-----DPSLLPQPNVQSSRP--LSEQLDGLTSLPDLKMTDGA 559
 91 YL-----PELGKSEMEFTESKPRLELLPDLLRPSLLATQPAQSGEDLTDRSLTGK 144
 560 ILGLKYKIPELGKQVEDL-----FTAVLSPANT----- 588
 145 LEVSPSPSPHTPELELPPRLSKBELIONMDRVDREITMVEQQLSKLKKQQLLEEA 204
 589 -QTPPLPQPP---PPTQLLP-----IHNQAFSR-MPLMNGLI-----G 622
 205 AKPPEKPVSPPIESKHSRLVQIYIDENRKAEEAAHRLGLEGLPQVELPLYNQPSDTR 264
 623 SSPHLPHNSLPPGSLGTFSIAQSSYPDARDKNSAFN-----PMASDPNNS- 669

QY 265 QYHENIKINQAMRKKLILYFKERNHARKQWKQKFCQRYDQLEALEKKYVERIENRRRA 324
 DB 670 -----WTSS-----APTVEG-ENDTMSNA 687
 QY 325 KESKVEYYEYKOPPEIRKORELOERMQSRVQGRGSLSMSAARSEHEVSEIIDLSEQEN 384
 DB 688 QRSTLK--WEK----- 696
 QY 385 LEKOMROLAVIPMLYDADQRIKFINNGLMADPMKVYKDOVMNMWSEOEKETPREKF 444
 DB 697 -EALGEMATVAPVLY---TWINFNLKEEPPD--WTRVKQIAKLWRK----- 739
 QY 445 MQHPKNFGLIASFLERKTVACVLYYYLTCKENYKSLVRRSYRRRKSGKQQQQQQQQQQ 504
 DB 740 -----ASSQERAP-----YVQKARDN-----RAALRINKVQMSNDKMKRQQ 775
 QY 505 QQQQQQPMRPSRQOEKDEKEKEKEKEKEKEKEKEKEKEDELLKEKTDGDDGDEKEAV 564
 DB 776 QQDSIDPSSRIDSELFQKPLKORESEHEQ----- 805
 QY 565 ASKGRKTANSQGRKGRGRI--TRSMANEANSEBAITPQSAELASLMELNESSRTEEMET 622
 DB 806 -WKFRQOMRQKQKQAKIEATQKLEQVKNQEQ---QQQQQFQSGHL-----LVQSGSDI 856
 QY 623 AKKGLLEHGRNWSAIARWVGSKTVSQCKNFYFNKKRQNLDEILQOHLKMKERNARRK 682
 DB 857 PSSGI-----QSPLTPQFGNGMSPAQSF-----HKLFTKQ----- 888
 QY 683 KKKAPAAASEEAAAF-----PPV-----VEDEMEASGVSGNEEMVEEAALHASGN 729
 DB 889 PSTPTSTSSDDVFKVPQAPPPPPAPSRIPIDLSLQA----- 926
 QY 730 EYPRGCSGPAVNNSDDTESIPSHTEAAKOTGQNGKPPATLGADGPPPGPTTP--- 786
 DB 927 -----QTSQPPSPQVFS---PGSSNSRPPSPMDPYAKNVGTPRPPVVG 966
 QY 787 ---RRTSRAPTEP-TPASEA-----TGAPTPPPAPP-----SPSAPPVVPVK 824
 DB 967 HFSRNSRAAPVENCCTPLSSVRPLQMNETTANRPSFVRDLCSSTTNDNDPYAKPDP- 1025
 QY 825 EKEEETAAAPVVEEGEOKPAAAEAVDTGKAEFPVKSECTEEAEGBGAKGDAEAAE 884
 DB 1026 -----RPVMTDQFKSLGLSRSPVVEQT-----AKG-----PIA 1056
 QY 885 ATAEGALKAEKEGGSGRATTAKSCAPQDSSSATSADAEVDEAGGDKNRLLSRPSL 944
 DB 1057 GTSDFHTKP-----SPRADVFQQRIP-DSYARPLLTAPLDSGGPGFKTP-MQPPSS 1108
 QY 945 LPTGDPANASPKQLDLKQLKQRAAAIPIQVTKVHEPPRDEAAAPTAPAPPAPPPQN 1004
 DB 1109 QDPYGSV-SQASRRLSVD---PYERPALTFRPINDFNHNSQNDPYSPQLTTPHAVNES 1163
 QY 1005 LOPESDAPQPGSSPRGKSR-----SPAPPADKEFAAAEQKLPGDPPCWTGSLPFP 1056
 DB 1164 FAHPSRAFSQPGTISRPTSDQDPYSGPPGTPRPVDSYSQSSGSGTARGNTDPSYSGPTPR 1223
 QY 1057 --VPPREVTKASPHADPSAFS-----YA-PPGHPLP-LGLHDTARVLP 1097
 DB 1224 TTVDP---YSQQFQTPRPTQTDLFVTPVYNQRHSDPYAHPPTGTPRGISVPSYQPPATP 1280
 QY 1098 RP-----PTI-SNPPPLISSAKH-----PSVLERQIGAISQMSVOLHVPYS 1138
 DB 1281 RPRISGFTSRSSRTRVLMFNQDPFLQAAQNRGPAALPGPLVRPDPDTCQ----- 1329
 QY 1139 EHAKAPVGVMTGL-----FLPMDPKKLAPFSGVKQ-----EQLSPR 1175
 DB 1330 --TPRPPGP---GLSDTFESRVSFAARDPYDQSPMTPRSQSDSFGTSQTAHDVADQPRG 1384
 QY 1176 GQAGPPESLGVPQAQASVLRGTALGSVFGGSGITKIPST----- 1215
 DB 1385 SEGSCFCASSNPMHSQQQFSG--VSQPLPGVPPTSGVTDTQNTVMAQADTEKLQRQKL 1442

Qy	1216	----	RVPSDS-AITYKGSITHGTPADV-----LYGTTTRII	1244
Db	1443	REIILOQOQKCIAGROEKGSQSDSPAVPHPGPLQHWQENPVNAOFRPPPPYPCGNIR---	1499	
Qy	1248	GEOSPSRLDRG-REDSIPKGGH-----VIYEKKKGH--VLSYEGGMSVTQCSKEDGRSS	1297	
Db	1500	---SPVAPPLGPRYAVFPKQORGYPDPDVASMGMRPHGFRFGPGSGHGTMPQOE---RF	1553	
Qy	1298	SGPPHET---AAPKR-----TYDMEGRVGRAISSASIEGLMGRAIIPPERHSHPHLK-EQ	1348	
Db	1554	LVPPQOIQSGSVSPOLRRSVSDM-----PRPLNNSQMNPNVGL---PQHFSQSPLPVQO	1605	
Qy	1349	HHIRGSI-----TQGI PRSYVAQEDYLRREAKLKKREGTPPP---PPPS	1390	
Db	1606	HNILGOAYIELHRAPDGRQRLPSPAPGSGVWEASSNL--RHGNFIIPRPDPFGRHUTDPM	1663	
Qy	1391	RDLTEAYKTQALGFLKPKAHEGLVATVKEAGSIH--EIPREELRHT--PELPLAPRPL	1446	
Db	1664	RRPPQGLPNQ----LPVHPDLLEGPPSQOQGHSHVSSMMVTMLNHLPLGGEFSEAPLST	1719	
Qy	1447	KEGSIOTGTPLYKDYDTGASTTGSKKHDSVRSLSIGSPGRTPPVHPVLDMVADARALE-RACYE	1505	
Db	1720	SVFSETSDNLQITQSDGLEGKELDSD-----PSVKELDV-KOLEGEVEXDOLD	1769	
Qy	1506	ESLKSREPTASSSGSIAR-----GAPVIV---PELGKPRQSPL	1541	
Db	1770	EDLENL--NLDTEGKVVELDTLONLETNDPDLDDLRSGEFDIIAYTDFELDMGDKSM	1827	
Qy	1542	TYEDHGAPFAGHLPGRGSPVTRBPTPKLQGS--LSSSKASQDRKLTST-----	1588	
Db	1828	FNEELDLPIDDKL--DNQCYSVEPKKEQENKTLVLSDKHSPKCKSTVTNEVKTEVLSPN	1885	
Qy	1589	-----PREIAKSPHSTVEHPH-----HPISP--YEHLLR-----	1616	
Db	1886	SKVESKETENKDNKNDVTPCSQAHAHSDLDNGEKTSLHPCDDPLFEKRTNRETAGPS	1945	
Qy	1617	-----GVSGVDLYRSHIPLA---FDPSTI-PRGIPLDAAAAAYLPR	1653	
Db	1946	ANVIQASTQLPAQDVINSGITGSTPVLSSL-LANEKSDNSDIRPSGSP-----	1993	
Qy	1654	HLAPNTPYP-----HL--YPPYLIRGVPDTAALENROTIINDYIT-SQMHHTATAMAQ	1705	
Db	1994	---PPPTLPASPSNHVSSLPFFIA---PPGRVLDN---AMNSNVTVSVRNH---VFSG	2040	
Qy	1706	RADMLRGLSPRESSALANYAAGPRGIIIDLQOVPHLPVLVPTPGTATAMDRLAYLPTAP	1765	
Db	2041	GVQNPGLIPEQST--VNHSLG-----TKGPAT-----QTGP	2070	
Qy	1766	QPFSSRRSSPLSPGPGTHLTKPTTTSSSERER-----DRDRERDRDR	1810	
Db	2071	Q--TSQSGTSSMS--GPQOLMIPTQLAQONRERPLLLEEQPLLLODLLDOERQEQOQORQ	2126	
Qy	1811	EKSILTSTT-----TVBHAPIWRPGTQS-----	1834	
Db	2127	MQAMIRORSEPPFFNIDFDAITDIPIMAKMVALKINKVMAQNNLGMPPMVMSRFPFMQO	2186	
Qy	1835	--SGSSSGSGGGSSSRPASHAHQHSPLSP-----RTQDAL	1871	
Db	2187	VVTGTQNSEGNLGPQAIPOGSIITHQISRNPNNPFGFVNDQSQRKQYEWI-OETQQLL	2246	
Qy	1872	QORPSVL-----HNTGMKGIIITAVEPSKPTVLJRSTSSPVRPAATFFPATHCHPLGGTL	1925	
Db	2247	QMQQYLEEQIGARHKSFKAL-----SAKORTAKKAGREFPEDEAQLKHVTE-----	2294	
Qy	1926	DGVYPTLMEPVLLPKEAPRVARPERPADTGHAPLAK-----PPA-----RSGLE	1970	
Db	2295	-----QGSWVGKOLEQIRKQOKEHAELIEDYRIKQOQCCAMAPMTMPSVQPQPPLI	2346	
Qy	1971	PASSPSKGSER-PLVPP-----VSGHATIARTPA-----KNLAEPHASPPPP--	2012	
Db	2347	PGATPPTMSQPTFFMPVPOQLQHOHQHTTVISGHTSPVRMPSLPGQWQPNPSAHLPLNPPRI	2406	
Qy	2013	-----APPASASDPREHKTOSKPFs-----	2032	

Db	2407	QPPIAQLPIKTCPPAGTYSNAPQSGPPRPVFEFDDNPFSEFQBRERKERLREQQERQ	2466
Qy	2033	-----IQELERLSLGYHGSSVSP-----EGVBPVSPVSS	2061
Db	2467	RIGLQMEVDRQALQORMEWQHGMVGSIISSRTSVSQIPFYVSSDLPCDFMQPLGFLQQ	2526
Qy	2062	PSLTHDKGLPKHLEEL-----DKSHLEGEELRP-----KQGPVKLGGEAAHLP	2104
Db	2527	S-----PQHQQMGQVLQQQNIQQGSIINSPTQTMTNERRQVGPSPFVPDPSIP	2578
Qy	2105	HLRPLPES-----OPSSPLLQATPAGVKGHQVRVTLAQHISEVITQD	2146
Db	2579	VGSFNFSSVKQGHGNI-SGTSFQOSPVRPSPFTPALPAAPPV-----ANSLSLPGQD	2628
Qy	2147	YTRHHQQLSAPLAPLAPLYSPFGASCVPDLRLRRPPSDLYLPPDPDHGAPARGSPHSEGGKRS	2206
Db	2629	STITHG-----HSYFGSTQSLIQLY-----SDII-----PEEKKKKR	2661
Qy	2207	PEPNKTSVLGGGEGDIEPVSPPEGMT---EPGHSRSADVPLLYRDGEOTEPEPSRMGSKSPG	2263
Db	2662	TRKKRD--DDAESTKAPSTPHSDITAPPTGGISETTTPAVSTPSELQQADQESVEPV	2719
Qy	2264	NTSQPAPFSL--TEGNSAMVKSKEINKKLTHNRNEPEYNI SQPGTEIFNMPAITGT	2322
Db	2720	GPSTPNMAAQGLCTELEKNLP-----NSDFSQATFNQQTVANSEVDKLSMETPAKT--	2770
Qy	2323	GLMYSRQAVQEHASTNMGLEAIRKALMKCYDQWEESSPPLSANAFNPLNASASLPAAMP	2382
Db	2771	-----BEIKLEKAEATESCPG-----QEEPKLEBQNGSKVEGNA---VACP	2807
Qy	2383	ITAADGRSDHLLTSP---GGGKAKVSGRPSSRKAKSPAPGLASGRPPSVSVSHSEGDC	2439
Db	2808	VSSAQ-S-PPHISACAPAAKGDGSELKHLKLNKSSS-----LLNQKPE--GSICSEDDC	2859
Qy	2440	NRETPILTNRVWEDRPSAGSTPPFPYNPLIMRLQAGVMASPPPPLPAGS-----	2488
Db	2860	TKDNKL---VEKQNPAGEGLQT-----LGAQMGGFGCGNQLPKTDGGSETTKQRSKRQTQ	2910
Qy	2489	--CPLAGPH---HAWDEEPKPLLCQOYETLS	2514
Db	2911	RTGEKAAPRSKKRKKDEBEKQAMYSSTDFT	2941
RESULT 59			
ABP69736			
ID	ABP69736 standard; protein; 4025 AA.		
XX	ABP69736;		
XX	20-JAN-2003 (first entry)		
DT	Human polypeptide SEQ ID NO 1783.		
XX	Human; genome mapping; gene therapy; food supplement; virus; fungus;		
KW	cell-proliferative disorder; neurodegenerative disease; bacterial;		
KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;		
KW	multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;		
KW	arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;		
KW	antiparkinsonian; antidiabetic; immunosuppressive; dermatological;		
KW	haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;		
KW	antiarthritic.		
OS	Homo sapiens.		
XX	WO200270539-A2.		
XX	12-SEP-2002.		
XX	05-MAR-2002; 2002WO-US005095.		
PF	05-MAR-2001; 2001US-00799451.		
XX			
XX			

QY 1447 KEGSITQGTPLKYDTCASTTGSKKHIVRSILGSPGRTPFPVHPLDVMADARALE-RACYE 1505
Db 1726 SVPSSETSNLQITTOPSGLEBKLDSD-PSVKELDV-KDLEGVEVKDLD 1775
QY 1506 ESLKSRFGTASSGGSIAR-GAPVIV-PELGKPRQSP 1541
Db 1776 EDLENL-NLDTEDGKVVLELDTLNLNDLRLSSEFIIAYTDPDLMDGDKSM 1833
QY 1542 TYEDHGAAPAGHLPRGSPVMTREPTLRQGS-LSSKASQDRKLTST- 1588
Db 1834 FNEELDLPIDDKL-DNQCVSVBPKKEQENKTLVLSKHSQKSTVTNEVKTEVLSN 1891
QY 1589 PREIAKSPHSTVPEHP-HPISP-YEHLRL- 1616
Db 1892 SKVESKETEKENDKNDVTPCSQASAHSLDNGEKTSILHPCDPLFKRTNRETAGPS 1951
QY 1617 GVSQVDLYRSHIPLA-FDPTSI-PRGIFLDAAYLPR 1653
Db 1952 ANVIQASTQLPAQDVINSGITGSTVPLSSL-LANEKSDNSDIRPGSP- 1999
QY 1654 HLAENPTYP-HL-YPPYLIRGYPDTAALENQTIINDYIT-SQOMHNTATAMAQ 1705
Db 2000 PPTLPASPNSHVSSLPPFIA-PPGRVLDN-AMNSNVTVSVRNH-VFSQ 2046
QY 1706 RADMLRGLSPRESSLALNVAAGPRGIIDLSQVPHLPVLPPTPGTPTATAMRLAYLPTAP 1765
Db 2047 GUVQNFGLPGQT-VNHSLG-TGKPA- 2076
QY 1766 QPFSSRHSSPLSPGPGTHLTKPTTSSSER-DRDRDRDR 1810
Db 2077 Q-TSQSGTSSMS-GPQQLMIPTQLAQNRRERPLLEQLLQDLDQERQEQOQ 2132
QY 1811 EKSILSTT-TVEHAPIWEPGTEQ- 1834
Db 2133 MQAMIRQSEPFENIDFADITPIMKAKMVALKINKVMAQNGLGMPMWSRFFPMQ 2192
QY 1835 SGSSGSGSGGSSRSPASHAHQHSPI-RTQDAL 1871
Db 2193 VVTGTQNSEQNLGPQAIPODGSITHTQISRPNPNFCGFVNDQKQYEWLQETQQL 2252
QY 1872 QORPSVL-HNTGMKGIITAVBSKPTVLRTSTSPVRPAATFPFATHCPLGGTL 1925
Db 2253 QMOKYLEQIGAHRSKKAL-SAKORTAKAGREPEEDAEQLKHVTE- 2300
QY 1926 DGVPYTLMEVLLPKEAPRVARPERPADTGHAFK-PPA-RSGLE 1970
Db 2301 QSMVQKQLEQIRKQKQKHAELIEDYRIKQOQOCAMAPPTMPSVQOPPLI 2352
QY 1971 PASSPSKGSBR-PLVPP-VSGHATIAATPA-KNLAPHASPDPP- 2012
Db 2353 PGATPTMSQPTFMVQQLHQOHTTVISGHTSPVRMPSLPGWPNAPAHPLNPPRI 2412
QY 2013 APPASADPHREKTSQKPS- 2032
Db 2413 QPPIAQLPIKTCTPAPGTVSNANPQSGPPPRVFEFDDNNPFESFQERKERLEQOERQ 2472
QY 2033 IQELELSRIGYHGSYSP-EGVPEVPSVSS 2061
Db 2473 RIQLMQEVDQRALQORWEMEGHWGSEISSRTSVSQIPFYSSDLPCDFMPLGLQ 2532
QY 2062 PSLTHDKGLPKHLEL-DKSHLEGELRP-KQGPVKLOGEAHL 2104
Db 2533 S-FHQOQMQVQLQQNIQGSINSPTQFMQTNERRQVPPSPFVPSIP 2584
QY 2105 HLRPLPS-OPSSPLLOTAPGVKHQVRVTLAQHTSEVITOP 2146
Db 2585 VGSFNSVSKGHGNSLGSFQSGVPSRPSFTPALPAAPPV-ANSLPCQGD 2634
QY 2147 YTRHPQOLSAPLPAPLYSPGASCPVLDLRRPSPDLYLPPDHGAPARGSPHSEGGKRS 2206
Db 2635 STTHG-HSYFGSTQSLIQLY-SDII-PBEKGKKR 2667

QY 2207 PEPNKTSVLGGEDGIEPVSPPEGMT-EFGHRSASVYPLLYRDGEOTEPERMGSKSPG 2263
Db 2668 TRKKRD-DDAESTKAPSTPHSDITAPPTPGISETTSTPAVSTPSELPOADQESVEPV 2725
QY 2264 NTSQPPAFSKL-TESNSAMVSKKQEKINLKNTHNRNEPEYNISQPGTEIFNMPEAITCT 2322
Db 2726 GFSTENMAAGQLCTELEKLP-NSDFSQATPNQOITYANSEVDKLSMETPAKT- 2776
QY 2323 GLMVTYRSQAVQEHASTNMGLEAIRKALMGKYDOWEESPPLSANAFNPLNASASLPAAMP 2382
Db 2777 EEIKLEAETESCPG-QEFPKLEEQNGSKVEGNA-VACP 2813
QY 2383 ITAADGRSDHTLUTSP-GGGKAKVSGRPSRKAAPGLASGDRPPSVSVHSEGDC 2439
Db 2814 VSSAQSG-PPHSAGAPAAKGDGSGNELLKLLKNKSSS-LLNQKPE-GSICSEDDC 2865
QY 2440 NRTPLTNRVWEDRSPSSAGSTFPYFNPLMRLOAGVMASPPPPGGLPAGS- 2488
Db 2866 TKDNKL-VEKONPAEGLQTLGAQMGGFGCGNQLPFTDGGSETKKQSKRTQ 2916
QY 2489 GPLAGPH-HAWDEBPKPLLCQYETLS 2514
Db 2917 RTGEKAAPRSKKKDEEKKQAMYSTDTFT 2947
RESULT 60
ABB60291
ID ABB60291 standard; protein; 2951 AA.
XX AC ABB60291;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 7665.
XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL04394.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX PS Disclosure; SEQ ID NO 7665; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceuticals. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ffp.wipo.int/pub/published_pct_sequences


```
QY 1900 STSTSPVRPAATFFPAPHCPLGGTLDGVYPTLMPEVLLPKEAPRVAR-----PE 1949
Db 2073 -TQKETPLPQKQ-----PLSKVKD-----EPEKVNKEPKVPQKESQTKLKEEPE 2117
QY 1950 RPRADTGHAFKAPPAR-SGLEPASPSPK--GSEPRPLV-----PPV 1988
Db 2118 RVTKTTPQKEPRKEPLRQSEDEPFESPEFDEPLFMTKTHTTAEMKRQKDILNRPSV 2177
QY 1989 SGHATIAIAR-----TPAKNL-----APHHA 2007
Db 2178 FQORTPERKSTTSPSTKLNGTRGPRSPSTNLITEKRSYRQVTVNSKPGTRKTTPSAN 2237
QY 2008 SPDPAPPASADPHR-EKTSQKPFQI--ELESLGLVHGSSY---SPEGVEPVSPVSS 2061
Db 2238 SPAQSPPTTTSIKRMEQISQSQSVVQVDVDEVGVPAPPFISHSEKPGKSP-SPTSS 2296
QY 2062 PSLT-----HDKG 2069
Db 2297 RSLRSRSPSRPKSIITTTTNTGNVSRNRNVEFVHETHVDSEPTGRRRPSYMDHTKS 2356
QY 2070 LPKHLE---ELDKSHL-----EGELRPKQPG-----PVKLGGEAAHLPHLRPLP 2110
Db 2357 SLEHTRDSLEINKSHYSKSMEDDSPVEPRNPNSVKFDPVRKSSRGADPEPKTSLK 2416
QY 2111 ESQPS-----SPLQTPGVKGHQVVTLAQHISE-----VI 2143
Db 2417 GKDESDLELEETEIFDLQRLKLETVASYEMRRIRRAQMLIRKMNINAGTTTTITI 2476
QY 2144 TDYTRHHFQQLSAPLAPLYFPFGASCVPDLRRPPS-----2181
Db 2477 TTSTT-----PGKSLPKIRRDQSPAGAAEVKTKVTRTTTERRQQO 2518
QY 2182 -----DLYLPPDHGAPARGSPHSGGKRSPENKTSVLGGEGDIEPVSPPEGMTGP 2235
Db 2519 QRVEQVDSITPI-----APKTSFH-----GKPPVKPRERSASPAK---RHSPP-GKQSPG 2567
QY 2236 H-----SRSAVPLLYRD-----GEOTSPSRMG-SKFGNTSQPP 2269
Db 2568 DRSTTTTITKVTSTTRGAPSPKPAQGP-IWADRSKVLKGHAIVPQTNGTTPRKGSSTT 2626
QY 2270 AFPSKLTESNAVSKQKEINKLNTNRNPEYNISQPGTEIFNMPAITGLMTYRS 2329
Db 2627 SSGKITRTMTS--SSTTTSSSTTNRNKRQREDSI-----TSSYGVGPTDENGULPFGI 2680
QY 2330 QAVQEH-----STNMGLEALIRKA 2349
Db 2681 RALKKATPAEPCETKQVTVGVIEEQFYSDNKSPPRHRKELIYSSNADELAIAIKQ 2740
QY 2350 LMGKYDQWESPP-----SANAFNPLNASASLP 2378
Db 2741 LQ-----DEDDSPPLLDARVVREPKVESQQSLP 2770

RESULT 61
AAE21713
ID AAE21713 standard; protein; 2429 AA.
AC
AC AAE21713;
XX
XX
DT 16-JUL-2002 (first entry)
XX
DE Human PKIN-8 protein.
KW Human; kinase; enzyme; PKIN-8 protein; immune system disorder; anaemia;
KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;
KW asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
KW AIDS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; developmental disorder;
KW Down's syndrome; gene therapy; protein therapy; cytostatic.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH
```

```
FT Domain 374..522 /note="Protein kinase domain"
FT Domain 375..522 /note="Protein kinase domain"
FT Domain 376..649 /note="Eukaryotic protein kinase domain"
FT Domain 378..636 /note="Protein kinase domain"
FT Domain 379..522 /note="Protein kinase domain"
FT Domain 489..507 /note="Tyrosine kinase catalytic domain"
FT Domain 549..697 /note="Protein kinase domain"
FT Domain 570..592 /note="Tyrosine kinase catalytic domain"
FT Domain 946..1034 /note="PDZ domain"
XX
PN WO200218557-A2.
XX
XX 07-MAR-2002.
XX
XX 31-AUG-2001; 2001WO-US027219.
XX
XX 31-AUG-2000; 2000US-0229873P.
PR 08-SEP-2000; 2000US-0231357P.
PR 14-SEP-2000; 2000US-0232654P.
PR 22-SEP-2000; 2000US-0234902P.
PR 29-SEP-2000; 2000US-0238499P.
PR 06-OCT-2000; 2000US-0238389P.
PR 13-OCT-2000; 2000US-0240542P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Bandnan O, Nguyen DB, Walia NK, Hafalia AJA, Yao MG, Gandhi AR;
XX Gururajan R, Ding L, Patterson C, Yue H, Baughn MR, Tribouley CM;
XX Thornton M, Elliott VS, Lu Y, Ison CH, Au-Young J, Tang YT;
XX Azimzai Y, Burrill JD, Marcus GA, Zingler KA, Lu DAM, Lal PG;
XX Ramkumar J, Warren BA, Kearney L, Policky JL, Thangavelu K;
XX Burford N;
XX
XX WPI; 2002-329769/36.
DR N-PSDB; AAD344305.
XX
XX New human kinases, useful for diagnosing, treating or preventing immune
XX system disorders (e.g. Crohn's disease), neurological disorders (e.g.
XX epilepsy), or cell proliferative disorders (e.g. cancers such as leukemia
XX or lymphoma).
XX
XX Claim 63; Page 156-162; 218pp; English.
XX
XX The present invention relates to human kinases (PKIN) and polynucleotides
XX encoding such proteins. PKIN sequences of the invention are useful for
XX diagnosing, treating or preventing disorders associated with aberrant
XX expression of PKIN, particularly immune system disorders (e.g. acquired
XX immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease,
XX anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-
XX Tooth disease or seizures), cell proliferative disorders (e.g. cancers
XX such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma),
XX and developmental disorders (e.g. Down's syndrome). They are also used in
XX gene therapy and protein therapy. The present sequence is human PKIN-8
XX protein
XX
XX Sequence 2429 AA;
XX
XX Query Match 3.3%; Score 429.5; DB 5; Length 2429;
XX Best Local Similarity 19.9%; Pred. No. 7.8e-15;
XX Matches 551; Conservative 305; Mismatches 982; Indels 931; Gaps 139;
XX
XX 38 GLLEYQHH-----SRDYASHLSFGSI-----IQPQRRRPSLLSEFQPGNERSQELHLRPE 87
XX 215 GVLSFTTHQIIEIARDCLDKSHQGLTSRYFLELQHLDKLL-----QEADRSE 264
```


Db	1973	PS	TAEPSSSQDPPKPVAAHS	-----SSSHKPRPGDPGPKTKHPRSLSSQKP-	2023
Qy	2032	SI	-----QELELRSLG	---YHGSYSPEG---VEPVSPVS-----SPSLTHDKGLPKHLEEL	2077
Db	2024	SV	GATKKGEPATOSLGSSREGKHSKSGDPVPATPGSONKASDGGICQEGGSPVPLHT	2083	
Qy	2078	DK	SHLEGEIRPKQPG-PVKLGGEAAHLPHLRPLPESQSSPILLQTAPGVKGHQRVVTLA	2136	
Db	2084	DR	APLDAKQPTSGGRPLEVLEKPVHLPRPGHPGPSEPADQKL--SAVGEK	2132	
Qy	2137	QH	ISEVITQDTRHHQPOOLSAPLPALYSPFGASCVPDLR---RPPSDLYLPPDPHGAP	2193	
Db	2133	-----	QTLSPKHPK-----PSTVKDCP-TLCKQTDNRQTDKPSQ	2167	
Qy	2194	ARG	SPHSEGGKRSPE-----PNKTSVLGGEDGIEPVSPPGMTPEGHSRAVYPLLRYDG	2249	
Db	2168	AA	NTDRRAEGKCKTEALYAPAGDKLEAGLSFVHSENRLKGAERPAAGVGKGGFP	2221	
Qy	2250	EQ	TEPSRMGSKSGNTSQPPAFTSKLTESNAVYKSKQBINKKLANTHNEPEYNISQP	2309	
Db	2222	-----	EARGKFG-PQKPP-----TEAD-----KPNGMKRSP-----	2247	
Qy	2310	GTE	IFNPAITGCLMTYRSQAVQEH-----ASTNMGLEALIRKALMGKYDQWEEPPLSA	2365	
Db	2248	-----	SATGQSSFRSTALPEKSLSCSSSFPETTRAGVREASAASD-----	2287	
Qy	2366	NA	PNPLNASLPAAMPITAADGRSDHTLTSPGGGK--AKVSGRPSR-----	2412	
Db	2288	-----	TSSAKAAGMLELPAPNRDRKRAQPAGEGRTHMTKSDSUPSFVRVSTLPLESHH	2341	
Qy	2413	-----	-----KAKSPAPGLASGDRPPSVSVSHSGDCNRRTPLTN	2447	
Db	2342	PD	NTMGGASHRDALSVTATVGETKGPAPA-----QPPARKQNVGRDVTKPSPAPN	2396	
Qy	2448	RW	EDRPS	2456	
Db	2397	---	TDRLPS	2402	
RESULT 62					
AAB12000					
ID	AAB12000 standard; protein; 2091 AA.				
XX	AAB12000;				
DT	19-DEC-2000 (first entry)				
XX	Rat p3103 protein.				
XX	Rat; p3103 protein; Synamon; SH3 domain; PDZ domain; SAPA1 interaction;				
KW	neuronal function; nervous system; neurological disorder.				
OS	Rattus sp.				
XX	Key	Location/Qualifiers			
FT	Domain	554..613	/note= "SH3 domain"		
FT	Domain	654..749	/note= "PDZ domain"		
XX	JP2000184884-A.				
XX	04-JUL-2000.				
XX	16-NOV-1998; 98JP-00325657.				
XX	16-NOV-1998; 98JP-00325657.				
XX	(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.				
XX	WPI; 2000-567637/53.				
DR	N-PSDB; AAA62000.				

XX	PT	A protein Synanon.			
XX	PS	Claim 1; Page 5-10; 13pp; Japanese.			
XX	CC	This sequence represents rat p3103 protein, also referred to as Synanon in the title of the specification. The p3103 protein contains both an SH3 domain and a PDZ domain, and was identified as being able to interact with the protein SAPA1 via the yeast two-hybrid system. p3103 protein is therefore thought to be involved in neuronal function, and may be useful for the study of the human nervous system, and for the diagnosis, prevention and treatment of various neurological disorders			
XX	CC	Sequence 2091 AA;			
QY	Query Match	3.2%; Score 429; DB 3; Length 2091;			
QY	Best Local Similarity	20.8%; Pred. No. 6.9e-15;			
QY	Matches 384; Conservative 140; Mismatches 627; Indels 694; Gaps 84;				
QY	593	EEAI	--TPQSAEL--ASMELN	ESSRWTE--REMETAKKGLLEHGRNWSAIAARMVGSKTV	646
DB	754	DEAVHKKASQAKRLPPPAISLR	SKSMTELEWSPWKKIEYEQQAAPVSMKKRTV	813	
QY	647	SQCKNFYNYKRONLDEILOQHLKMEKERNARRKKKAPAAASEEAAFPVVEDEME	706		
DB	814	YQ-----MALNKLDEL-----	AAQQTIS-----	ASE	836
QY	707	ASGVSGNEEMVEEAEALHASGNEVPRGECGPATVNNSSDTESIPSPHTEAAKDTGONG	766		
DB	837	SPGPG-----LASLGHKRPKGFFA-----	TESSFPDHRS-----	QPS	870
QY	767	PKPPATLGADGPPGPPPTPRRTS-----	RAPTEPA-----SEATGAPTPPPAPPSPSA	817	
DB	871	YDRPSFL-----PPGGLMLRQKSGIAEDDRPYLAPAMKFSRLSVPGSEDIPIPPPTT	925		
QY	818	PPPVVPEEKEEBETAAPVVEEGEEQKPPAAELAVDTCKAEPEVKSECTEAEEGPAKG	877		
DB	926	SP-----EPY-----STPPA-----	937		
QY	878	KDAEAAEATAEGALKAEKEGSGRATTA-----	KSSGAPQDSSSATCSADEVDEAE	930	
DB	938	-----PSSGRLTSPRGPPNPSGGPLPASSPSPGFPSPDTR	978		
QY	931	GGDKNRLLSPRSLTPTGDPANASQKPLDLKQLKQRAAALPPIQVTKVHEPP-----	RE	987	
DB	979	GGREKSL-----YHSAALPPAH-----	HIPI		

Db 1323 AFTSFLPPR--PLVHPLTGKALDPASPLGLALAAERALKESSEGGTQPQPPRPSPRY 1380
Qy 1340 -----HSPH--HLKEQHIR-----GSGITGIPRSYVEAQED 1369
Db 1381 DAPPPTLHHSPHSFPHARHEPVLRLWGDPARRELGYRAGLGQEKALATASPPAARS 1440
Qy 1370 YLR-----EAKLLKRGTTGTPPPPSDLTEYKQALGPKLKAHEGLVATVKEAG 1422
Db 1441 LLHRLPPTAGVGPLLLQLG--PEPTPHFGVSKAMRTAA----- 1478
Qy 1423 RSIHETPRELHTRP-LPLAPRLKEGSIOTGTLKYDTGASTGSKKHVRSLSIGSPG 1481
Db 1479 -----PEE-----PERLPHVRFLN-----CQARP-----PPAGTGSSTED-----GPG 1514
Qy 1482 RTFPPVHPLDVMDARALERACYEESLKSRRPTASSGGSIARGAPVIVPELKGKQPSPL 1541
Db 1515 --VPPSPRRVL-----PTSPTSRCNEENGLLLV-----LPPAPS 1550
Qy 1542 TYEDGAPFAGHLPRGSPVTMREPTRLQEGSLSSKASQDRKLTSTPREIAKSPHSTVP 1601
Db 1551 VDVGDC-----EFLFAEPLPPLEFNSFEK-----PESPLT 1582
Qy 1602 EHHPIPSYEHLLRGVGDLYRSHIPLAFDPTSIPIGILDAARAYVLPRLHAPNPT 1661
Db 1583 PGPPHLP-----DPPS--PATPLPAA----- 1602
Qy 1662 PHLYPPYLIRGYPDTAALENROTIINDYITSOQMHNTATAMAQADMLRGLSPRESSLA 1721
Db 1603 ---PPPAVAAPPT--LDSTASLTSY-----DSEVA 1629
Qy 1722 LNYAGPRGIIDLSQVPHPLVLPVPTGTPATAMDRLAYLPTAPQPFSSRHSSSPLSPG 1781
Db 1630 TLTOGAPAAPGD-----PPAGPPAPA---ADAPPAPQ-----G 1661
Qy 1782 PHLTKPTTSSSERDRDREREKSLTSTTVEHAPVWRPTEQSSGSS 1841
Db 1662 P---DPPTGDSGIEVDSRSSDHPLE---TISSASTLSL-----SAEGGNGTGGVA 1709
Qy 1842 GGGGSSSRPASHAHQSPISPRDQALQORPSVLHNTGMKIITAVEPSKPTVL--- 1898
Db 1710 GGGAGVAS-----GTELLDTVAVLDAQFGSGTGPFPVPPQLMTPS 1752
Qy 1899 ----RSTSTSSVPRPAATFPATHCPLGTLGVYPTLMPEVLLPKAPRVARPERPRAD 1954
Db 1753 KLRGRALGTSGNLRPG-----PSGGLRDPVTPT-----SPTVS--VTGAGT 1791
Qy 1955 TGHAFIAPKAPSGLEPASPSKSGSPRLVPPVSGHATIARTPAKNLAPHASPPPPAP 2014
Db 1792 DGLLALSACFGPSTAGVAGP---VAVEPEVPVPLPA--ASSLPKLLPWEEGPGPPPP 1846
Qy 2015 --PASASDPHREKTSQ--KPFSTQELRLSLGYHGSSVPEGVPSVPSPLT--HDKG 2069
Db 1847 PLPGPLSQOASNALATVAKASISLSKLLQFGGSS--TAGGALPWARGSGGSTDSHGCG 1905
Qy 2070 ---LPKHLBELDKSHLEGLRKPQGVKVLGGEAAHLPHLRPLPESQPSSS----- 2117
Db 1906 ASVIPERTSSLQORLSED-----SQTSLL-----SKPSSSIFQNWPKPP 1945
Qy 2118 -PLLTAPGVKGHORVVTVAQHSIVITQDYTRHHPQQLSAPLAPLYSFPGASCPLDL 2176
Db 1946 LPLPLTSGSVSSSTAAAGATSPSASASASTR-HLQGVFEFMRPPL-----L 1992
Qy 2177 RPPPSDLYLPPDPHGAAPARGSPHSEGGKSPSPNKTSVLGGGSDGIEPVSPPEGMTEPGH 2236
Db 1993 RRAPSLLPASDH-----KVSPAPRPSL-----PILP----- 2021
Qy 2237 SRSAVYPLYRDEQTEP-----SRMGSKSPGNTSQPPAFFSKLT 2277
Db 2022 -SGPIYPLGLF-DIRSLQEGREARLTPLPSLCHHTPGVLGVLEE 2064

RESULT 63
AAM24322

ID AAM24322 standard; protein; 1299 AA.
AC AAM24322;
DT 12-OCT-2001 (first entry)
XX Human EST encoded protein SEQ ID NO: 1847.
DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX Homo sapiens.
OS WO200154477-A2.
PN 02-AUG-2001.
PD 25-JAN-2001; 2001WO-US002687.
PF 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617745.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
DR N-PSDB; AAH98981.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
PT Claim 20; Page 1198-1201; 1275pp; English.
PS The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention
XX Sequence 1299 AA;
SQ
Query Match 3.2%; Score 428; DB 4; Length 1299;
Best Local Similarity 20.8%; Pred. No. 4.2e-15;
Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;
Qy 476 NENYKSLVRSYRRKGSQ-----QQQQQQQQQQQQQQQPMRPSSEKDEKE 524
Db 126 SQTIKSTTKRSPKPNKKTKVIESEITEHSVENQSSSSSSSSSSSTIWKSS 185
Qy 525 KEKEAKEE-EKPEVENDKEDLLKEK-----TDDT-SGEDNDEKAVASKRKTANSQ 575
Db 186 KNSAANRELQKLVKDNKKNTKKKPTKPPVVDAGSLONGDPKVTTPDSTTTOHKN 245
Qy 576 GRRKGRITRSMANESBEAITP-QQSAELASNELNESSRWTEEMETAKKGLLEGRNW 634
Db 246 VSTSPKIT--TAKPINRPSLPNSDTSKETSLTVNKETTVETKETTTNKQSTDGKEK 303
Qy 635 SAIAKRVGSKTVSQCKNFYNYKGRQNLDEILQKHLMKE---KERNARRKKKAPAAAS 691
Db 304 TTSKAKETQSIKTSKDL-----APTSKVLAKPTPKAETTTKGPALATTPKEPTTPPK 356
Qy 692 EEAAPPPVVEDEMEASGVSGNEEEMVEAEALHASNEVPRGECGPATVNNSSDTESI 751
Db 357 EPASTTP-----PTTIKSAPTPKE 380

QY	752	PSPHTEAAKDTGQNGFKPPATLGGADGPPPGPTTPRRTSRAPISPTPASEAATGATPPP--	809
DB	381	PAPTITTSAPTTTKEPAFTTT-----KEPAFTTTPKEPATTTTKEBPATT-TKSAPTTPKE	434
QY	810	PAPTSAPSAPPVVPVPEEKEEBETAAPPVEEGEEKPPAAABELAVDVTGKAABPVKSECTEE	869
DB	435	PAPTTPKKPATTPKE---PAPTTPKEPTTTPKEBPATTKEPATTTTKEPATTTTKEPAT	485
QY	870	AEEGPAKGDAABAATAAGALKAEKKEGGSGRAATTAKSGAPQDSSTATCSADEVEA	929
DB	486	APKKPATTPKEPATTTTKEPATTTTKE-----PSPTTPKEPATTTTKSAPTITKEP----	537
QY	930	EGGDKNRLLSRPSSLLTPTGDPRANASPKQLDLKLQKQRAAAIPTTQIVKHHP-----P	985
DB	538	-----APTITSAPTTPKEP-----SPTTKEBPATTPP	565
QY	986	REDA-APTKPAPPAPPQNQLPESDAPQQPGSSPRGKSRSPAPPADKFAFAEAOKLP	1043
DB	566	KEPATTPPKPATTPKEPATTPPKPATTTTKKPATTPAKPEPATTPKETATTPPKLT	625
QY	1044	GDP-----CWTSGLPFPVPPREVIKASHPADPDPSAFSYAPPGHPLGLGHDTARPVLPRP	1099
DB	626	PTTTEKLAPTTPEKPATTPBELAPTTPEETPTT-----PEEPAPT-TPKAAAPNTPE	679
QY	1100	PTISNP-PPLISSAKHSVLBRQICAI SQGSVOLHVYPYSEHAKAP-----VGPTVMGLP	1153
DB	680	PAPTTPKEPATTPKEPA-----PTTPKETATTPPKGTATTLKEP	720
QY	1154	LPMDPKKLAPSGVKQGOLSPRGOGAPGESLG-----VPTAQEASVLRGTALGVSPEGSI	1210
DB	721	APTTPKKPAP-----KELAPTTTKEPTSTSDKPATTP-----KGT-----	758
QY	1211	GIPSTRVPSDAITYRSGITHGTADVLYKGTIRIIGEDSPSLDRGREDSLPKGHVIY	1270
DB	759	--PTT--PKEPAT-----TPKEPATTPPKGTATTLKEPATTPPKPAKELAP-----	804
QY	1271	EKGXGHLVSYEGGSMVTQCKEDGRSSGGPHETHA--APKRTYDMMEGRVGRAISSASIE	1328
DB	805	-----TTTKGPTSTT-----SDKPATTPKETATTPKEPATTPPKPATTPPETP--	850
QY	1329	GLMGRAIPP-----ERHSPHILKEQHIIRGSI TQGI PRSYVEAOEDYL RREAKLKREGTP-	1384
DB	851	-----PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAEPTPK	887
QY	1385	----PppppSRDLTEAYKTOALGPLKLPAHEGVATVWKGAGRSIH EIPREELRHTEP-L	1439
DB	888	ALENSPKPEGVPTTKT-----PAATKPE--MTTAKD-----KTTERDLTPTETT	931
QY	1440	PLAPRLPKESGIT-----OQTPLKYDTGAS-----TTGSKK	1470
DB	932	TAAAPDKMETATTTTEKTESKITATTTQVSTTTTQDTPFKITTLTKTTLAPKVTTTTKT	991
QY	1471	HDRVSLIGSPORTFPPVPHLDVMADARALBRACVYESLKRSPGTASSGGG-IARGAPVI	1529
DB	992	ITITEINNKPEET-----APKQRATNSKATTPKPKPTKPKPTSTKKPKT	1039
QY	1530	VPELGKPRQSPLTYVEDHGAPFHAGHLPRGSPVTMRE--PTPRLOEGSLSS--KASQDRK	1584
DB	1040	MPRVRKPKTTP-----TPRKMTSTMPELNPTSKIAEAMLTTRPNQTPNSK	1086
QY	1585	LT-----STPREIAKSPHSTVPEHHPH-----PISPYEHLRGVSGVDLYRS	1626
DB	1087	LVEVNPKEBAGGAEGTTPMLLR-PHFVFWPEVTPMDYLPFRVFNQGIIN-----	1136
QY	1627	HIPLAFDPTSIPIRGIPLD	1644
DB	1137	--PMLSDETNICNGKPYD	1152

RESULT 64
 AAR26049
 ID AAR26049 standard; protein; 1404 AA.

each encoding a different MSF protein. Exons V and VI are thought to be related to the activity of the factor and are also implicated in the stability, folding and processing of the molecule. These exons are also thought to play a role in the observed synergy of MSF with other cytokines. Exons V - XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor, ie one or more of these exons may contain sequences which direct proteolytic cleavage, adhesion, organisation of the cellular matrix or extracellular matrix processing. Both naturally occurring and non-naturally occurring MSF's may be characterised by various combinations of alternatively spliced exons from this sequence, with the exons spliced together in differing orders to form different members of the MSF family. (Updated on 25-MAR-2003 to correct PN field.)

QY	Sequence 1404 AA;	
QY	Query Match	3.2%; Score 428; DB 2; Length 1404;
QY	Best Local Similarity	20.8%; Pred. No. 4.6e-15;
QY	Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;	
QY	476 NENYKSLVRRSRRRGKSKQ-----QQQQQQQQQQQQQQQMPRSPRSEKDEKE 524	
DB	126 SQTIKSTTKSRPKPPNKKTKKVIIEEITEHSVSENOESSSSSSSSSTIMKIKSS 185	
QY	525 KKEAEKE-EEKPEVENDKEDLLKEK-----TDDT--SGEDNDEKAEVASKRKTANSQ 575	
DB	186 KNSAANRELQKLLKVDKNKNRTKKTKPKPVVDGAGSLDNGDFKVTTPDTSTTQHNK 245	
QY	576 GRRKGRTSMANEANSEAITP-QQGAELASMELESSESWTEEMETAKGLLEHGRNW 634	
DB	246 VTSPPKIT--TAKPINRPSLSPNSDTSKETSILVUNKETTVETKTTTNNKOTSTDGKEK 303	
QY	635 SAIAIRMGSKTVSQCKNFYNYKKRQNLDEILQOHLKME---KERNARRKKKAPAAAS 691	
DB	304 TTSAKETQSIKTSADKL-----APTSKVLAKPTPKAETTTKGPAULTTPKETPTTPK 356	
QY	692 EBAAPFPVVEDEMEASGVSGNEEMVEEAEALHASGNEVPRGECGPGATVNNSDTSI 751	
DB	357 EPASTTP-----KEPT-----KEPT-----PTTIKSAPTTPKE 380	
QY	752 PSPHTEAAKDTGONGPKPATLGDGPPPGPTTPPRTSRAPLEPTPASEATGAPTTP-- 809	
DB	381 PAPTITKSAPTTPKEPAPTTP-----KEPAPTTPKEPAPTTPKEPAPTTP--TKSAPTTPKE 434	
QY	810 PAPPSPAPPPVVPVVEKEEETAAAPVVEEGEEOKPPAAEELAVDTGKAEPPVKSECTEE 869	
DB	435 PAPTTPKKPAPTTPKE--PAPTTPKEPTTPPKPAPTTPKEPAPTTPKEPAPTTP-- 485	
QY	870 AEGPAPKGDAEAAEATAEGALKAEKKEGGSGRAITAKSSGAPQDSDSATCSADEVDEA 929	
DB	486 APKKPAPTTPKEPAPTTPKEPAPTTPKE-----PSPTTPKEPAPTITKSAPTITKEP----- 537	
QY	930 EGGDKNLLSPRPSLLTPTGDPANASPKPLDLKQLKQRAAAIPIQVTKVHEP---P 985	
DB	538 -----APTITKSAPTTPKEP-----SPTITKEPAPTTP 565	
QY	986 REDA---APTAPAPPPQNLOPESDAPQPGSGSPRCKSRPAPADKEAFAEAOKLP 1043	
DB	566 KEPAPTTPKKPAPTTPKEPAPTTPKEPAPTITTKPAPTAPKEPAPTTPKETAPTTPKLT 625	
QY	1044 GPPP---CWTGSLPFPVPPREVIVKASHPADPSAFSYPAGHPLPLGLHDRTARVLP 1099	
DB	626 PTTPEKLAPTTPKEPAPTTPPELAPTTPPEPTPTT---PEEPAPTTPKAAAPNTPKE 679	
QY	1100 PTISNP-PPLISSAKHPSVLRIQIGAISQMSVQLHVPVSEHAKP-----VGPVTMGLP 1153	
DB	680 PAPTTPKEPAPTTPKEPA-----PTTPKETAPTTPKGTAPTTPK 720	
QY	1154 LPMDPKPLAPFSGVQKQLSPRQAGPPESLG---VPTAQEASVLGTLGSLVPGSGITK 1210	
DB	721 APTTPKKAP-----KELAPTTPKETSTTSDKPAPTTP-----KGTA----- 758	
QY	1211 GIPSTRVPSDAITYRGSITHGTPADVLKGTITRIIGEDSPSRUDRGREDSLPKGHVITY 1270	

DB	759 --PTT--PKBPAPT-----TPKEPAPTTPKGTAPTTPKBPAPTTPKBPAPKELAP----- 804	
QY	1271 EGKKGHVLISYEGGMSVTQCSKEDGRSSGPPHETA--APKRTYDMMEGRVGAISSASIE 1328	
DB	805 -----TTTKGPTSTT-----SDKPAPTTPKETAPTTPKEPAPTTPKBPAPTTPETP-- 850	
QY	1329 GLMGRAIPP---ERHSPHLLKQHHIRGSIITOGIPRSYVEAOEDYLRRBAKLLKREGTP- 1384	
DB	851 -----PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAEPKPK 887	
QY	1385 ----PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAEPKPK 887	
DB	888 ALENSPKPGVPTTKT-----PAATKPE-----MTTAKD-----KTTERRDLRTTPTT 931	
QY	1440 PLAPRLKESGIT-----QGTPLKYDTGAS-----TTGSKK 1470	
DB	932 TAAPKMTKETAITTEKTTESKITAITTQVTSITTTQDTTFFKITLTKTLTAPKVTITTKT 991	
QY	1471 HDVRLSIGSPGTTPFPVHPLDVNADARALERCYSELSKSRPGTASSSGS-IARGAPVI 1529	
DB	992 ITTTEIMNKPEET-----AKPKDRATNSKATTPKPKPTKAPKPTSTTKPKPT 1039	
QY	1530 VPGLCKPROSLTYEDHGAPFAGHLPRGSPVTMRE--PTPRLQEGSLSS---KASQDRK 1584	
DB	1040 MPVRVRKPKTTP-----TPRKWTSTMPLELNPFSRIAEAMLQTTTTRNPTNSK 1086	
QY	1585 LT-----STPREIAKSPHSTVPEHHPH-----PISPYEHLRLRGVGVLDYRS 1626	
DB	1087 LVEVNPKSEDAGAGETPHMLLR-PHVMPPEVTPDMDLPRVNPQGIIN----- 1136	
QY	1627 HIPLAFDPTSPRGIPLD 1644	
DB	1137 --PMLSDETINICNGKPPVD 1152	
RESULT 65		
AAAB29773		
ID	AAAB29773 standard; protein; 1404 AA.	
XX	AAAB29773;	
XX	28-FEB-2001 (first entry)	
XX	Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.	
DE	Human MSF; megakaryocyte stimulating factor; tribonectin; alternative splicing; joint boundary lubricant; O-linked oligosaccharide; osteoarthritis; tribosupplementation; tissue adhesion inhibition; friction coefficient reduction; gene therapy; antiarthritic; osteopathic.	
OS	Homo sapiens.	
XX	WO2000064930-A2.	
PN	02-NOV-2000.	
XX	24-APR-2000; 2000WO-USO10953.	
XX	23-APR-1999; 99US-002989970.	
XX	(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.	
XX	Jay GD;	
XX	WPI; 2001-024673/03.	
DR	N-PSDB; AAC81498.	
XX	Novel tribonectin polypeptide useful as lubricant for treating osteoarthritis, comprises O-linked lubricating moiety.	
PS	Claim 3; Page 7; 47pp; English.	

XX WPI; 2001-182721/18.

XX New composition comprising the camptodactyly-arthropathy-coxa vara-

PT pericarditis protein in combination with an anesthetic, useful for

PT treating osteoarthritis, or as lubricants of tissue and joints.

XX Example 1; Page: 34pp; English.

XX The invention relates to a method of treating osteoarthritis via the

CC administration of a composition comprising the camptodactyly-arthropathy-

CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.

CC The composition may further comprise a local anesthetic. The composition

CC of the invention may be administered via intra-articular or intravenous

CC injection. The human CACP protein is identified in the invention as being

CC megakaryocyte stimulating factor (MSF). The gene encoding CACP protein

CC (MSF) is located on chromosome 1q25-31, and mutations in this gene are

CC responsible for the heritable disorder camptodactyly-arthropathy-coxa

CC vara-pericarditis, in which patients have synovial hyperplasia without

CC evidence of inflammation. CACP protein (MSF) acts as a synovium

CC lubricant, and can be used to lubricate tissue and joints in the

CC treatment of osteoarthritis. The composition may be applied to reduce the

CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement

CC or joint damage). The present sequence represents human megakaryocyte

CC stimulating factor (MSF, CACP protein). Note: This sequence is not given

CC in its entirety in figure 4 of the specification, although a GenBank

CC accession number was given. This sequence was therefore obtained from

CC GenBank (U70316)

XX Sequence 1404 AA;

QY Query Match 3.2%; Score 428; DB 4; Length 1404;

DB Best Local Similarity 20.8%; Pred. No. 4.6e-15;

Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;

QY 476 NENYKSLVRSRRRQKSO-----QQQQQQQQQQQQQQQQQQMPRRSQEKEKE 524

DB 126 SQTIKSTTKRSPKPNKTKKTVIESEITEHSVSENQESSSSSSSTIWIKISS 185

QY 525 KEKEAKE-EEKPEVNDKEDLLKEK-----TDDT-SGEDNDEKAEVASKRGTANSQ 575

DB 186 KNSAANRELOKLLKVDKNKNTKTKPKPPVVDAGSLONGDPKVTPTDTSITQHNK 245

QY 576 GRRGRITRSMANEANSEBAITP-QQSAELASMELESSEWTEEMETAKKGLLEHGRNW 634

DB 246 VSTSPKIT--TAKPNRPSLPNSDTSKETSILTNNKETTVEKTTTNNKQSTDGKEK 303

QY 635 SAIAKMGKTVSQCNFVNYKKRQNLDEILQOHLKWE---KERNARKKKKAPAAAS 691

DB 304 TTSKAKETQSIKTSKDL-----APTSKVLAKTPKAEITTTKGPALTTPKEPTPTPK 356

QY 692 EEAAPPPVVEDEMEASGVSGNEEWEAEALHASGNEVPRGCSGPATVNSSDTESI 751

DB 357 EFASTTP-----KEPT-----PTTIKSAPTTPKE 380

QY 752 PSPHTAAKDTGONGPKPATLGDGPPPGPPTPRRTSRAPTEPTPASEATGAPTTP-- 809

DB 381 PARTTKSAPTTPKEAPITTT-----KEPATTPKEPATTTTKEPATTT-TKSAPTTPKE 434

QY 810 PAPPSPAPPPVVKKEKEEBEETAAPPVBEGBEQPPAAAEELAVDTGKAEBPVKSECTEE 869

DB 435 PAPPTPKKAPATTPKE--PAPTTPKEPTTTPKEPATTTKEPATTTKEPATTP----- 485

QY 870 AEEGPAGKGDAAEAATAGALKAEKKGSGRATTAKSSGAPQSDSSATCSADEVDEA 929

DB 486 APKAPATTPKGPATTPKEPATTTTKE-----PSPTTPKEPATTTKSAPTTPKE----- 537

QY 930 EGGDKNRLSPRLTPTGDRANASPOKPLDLKOLKORAAAIPIQVTKVHEP-----P 985

DB 538 -----APTTKSAPTTPKE-----SPTTKKEPATTP 565

QY 986 REDA--APTKPAPPAPPPONLOPESDAFOQPGSSPRGKSRSPAPPADKEAFABAKLUP 1043

DB 566 KEPAPTTPKKPAPTTPKEPATTTTPKEPATTTTKKPAPTAPKAPAPTTPKEPATTPPKKULT 625

QY 1044 GPPP-----CWTSGLPFPVPPPREVIVIKASPHAPDPSAFSVAPPGHPLPLGLHLDARVLP 1099

DB 626 PTTPEKLAPTTPKEPATTPPELATPTTEPTPTT-----PEEPAPT-TPKAAAPNTPKE 679

QY 1100 PTISNP-PPLISSAKHPSVLERIQIGAISQGSMSVQLHVPVSEHAKAP-----VGPVTMGLP 1153

DB 680 PAPTTPKEPATTPKEPA-----PTTPKETAPTTPKGTAPTTLKEP 720

QY 1154 LPWDKPKLAPSGVQEQOLSPRGOAGPPESLG---VPTAQEASVLRGALGVSFGGSIYK 1210

DB 721 APTTPKKPAP-----KELAPTTTKEPTSTTSKDPAPTTP-----KGTA----- 758

QY 1211 GIPSTRVPSDSAITVYRGSITHTGTPADVLYKGTITRIIGSDSPSRIDRGREDSLPKGHVY 1270

DB 759 --PTT--PKEAPT-----TPKEPATTPKGTAPTTLKEPATTPPKKAPKELAP----- 804

QY 1271 EGKKGHVLSEGGMSVTQCSKEDGSSSGPPHETA--APKRTYDMMEGRVGRAISSASIE 1328

DB 805 -----TTTKGPTSTT-----SDKPAPTTPKETAPTTPKEPATTTTPKPAPTTPETP-- 850

QY 1329 GLMGRAIP-----ERHSPHHLKEQHHRGSIOTGIPRSYVEAQEDYLRRKAKLLKREGTP- 1384

DB 851 -----PPTTSEVSTPTTTPKEPTTIHKSPDESTPE-----LSAETPK 887

QY 1385 ----PPPPSRDLTEAYKTQALGPLKLPKPAHGLVATVKEAGRSIHEIPREELRHTEP-L 1439

DB 888 ALENSPKFEGVPTTKT-----PAATKPE---MTTAKD-----KITERDLRTTPTT 931

QY 1440 PLAPRLKEGSIT-----QGTPLKYDTGAS-----TTGSKK 1470

DB 932 TAAPKMTKETATTTKTESKITATTTQVSTTTQDTPPKITTLTKTLTTLAPKVTTKKT 991

QY 1471 HDVRSLSGPGTFFPVHPLDVMADARALERACYESLSKSRPGTSSSSSGS-IARGAPVI 1529

DB 992 ITTITMKNPEET-----AKPKORATNSKATTPKPKPTKPKPTSTTKKPKT 1039

QY 1530 VPBLGKPRQSPLTYEDHGAPFAGHLPRGSPVTMRE--PTPRLOEGSLSSS---KASQDRK 1584

DB 1040 MPVRVKPKTTP-----TPRKMSTMPBELNPTSRIAEAMLOTTRPNQTPNSK 1086

QY 1585 LT-----STPREIAKSPHSTVPEHHPH-----PISPYEHLRLRGVSGVDLYRS 1626

DB 1087 LVEVNPKESEDAGAGETPHMLLR-PHVFMPEVTPDMDYLPVNPQGIIN----- 1136

QY 1627 HTPADFPTSIPRGIPLD 1644

DB 1137 --PMLSDETNICNGKPEVD 1152

RESULT 67

AAU32262

ID AAU32262 standard; protein; 1415 AA.

XX AC AAU32262;

XX DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #2753.

XX KW Human; vaccination; gene therapy; nutritional supplement;

XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US008656.

PR	18-APR-2000; 2000US-00552929.	QY	986	REDA--APTKPAPPAPPPQNLQPSDAPQOQSPGRKSRSPAPADKEAFAAQAQLP	1043
PR	26-JAN-2001; 2001US-00770160.	Db	577	KEPAPTTPKKPAFTTKKEPAPTTTKKPAFTAPKEPAPTTKRETAFTTKKLT	636
PA	(HYSE-) HYSEQ INC.	QY	1044	GDPP---CWTSGLPFPVPPREVIVIKASHPADPSAFSVAPPCHPLPLGLHDTARVLPVPP	1099
PI	Tang YT, Liu C, Drmanac RT;	Db	637	PTTPEKLAFTTPEKPAFTTPELAPTTPEEPTPTT-----PEEPAPT-TPKAAAPNTPKE	690
PR	WPI; 2001-611725/70.	QY	1100	PTISNP-PPLISSAKHPSVLERQIGAISQGSVQLHVPVSEHAKAP-----VGPVTMGLP	1153
PR	Nucleic acids encoding a range of human polypeptides, useful in genetic	Db	691	PAPTTKKEPAPTTKPEPA-----PTTKETAFTTKGTAFTTLKBP	731
PT	vaccination, testing and therapy.	QY	1154	LPWDKKLAPSGVKQEQOLSPRQAGPPESLG---VPTAQEASVLRGALGSPVGGSIK	1210
PS	Claim 20; Page 573; 765pp; English.	Db	732	APTTKKAP-----KELAPTTTKEPTSTTSKPAFTT-----KGTA-----	769
CC	The invention relates to novel human secreted polypeptides. The	QY	1211	GIPSTRVPSDSAITYRGSITHTGTPADVLVKGITRIIGEDSPSRDLDRGREDSLPKGHVY	1270
CC	polypeptides and antibodies to the polypeptides are useful for	Db	770	--PTT--PKEPAPT-----TPKEPAPTTKGTAPTTLKEPAPTTKPKAPKELAP----	815
CC	determining the presence of or predisposition to a disease associated	QY	1271	EGKKGHVLSYEGGMSVTCQSKEDGRSSSGPPHETA--APKTYDMMEGRVGRASIASIE	1328
CC	with altered levels of polypeptide. The polypeptides are also useful for	Db	816	-----TTTKGPTSTT-----SDKPAFTTTPKETAFTTKKEPAPTTKPKAPTTPETP--	861
CC	identifying agents (agonists and antagonists) that bind to them. Cells	QY	1329	GLMGRAIPP---ERHSPHLLKEQHHRGSIHQGIPRSYVEAQEDYLRREAKLLKREGTP-	1384
CC	expressing the proteins are useful for identifying a therapeutic agent	Db	862	-----PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAEPKPK	898
CC	for use in treatment of a pathology related to aberrant expression or	QY	1385	-----PPPPPSRDLTAYKTOALGPLKLPKPAHEGLVATVKEAGRSIHEIPRELRHTEP-L	1439
CC	physiological interactions of the polypeptide. Vectors comprising the	Db	899	ALENSPKBPGVPTTKT-----PAATKPE---MTTAKD-----KITERDLRTTPEIT	942
CC	nucleic acids encoding the polypeptides and cells genetically engineered	QY	1440	PLAPRPLKEGSIT-----OGTPLKYDTCAS-----TTGSKK	1470
CC	to express them are also useful for producing the proteins. The proteins	Db	943	TAAPKMTKETATTTKTTESKITATTTQVSTTTQDTTPTTKITLTKTTTLAPKVTTKT	1002
CC	are useful in genetic vaccination, testing and therapy, and can be used	QY	1471	HDVRSLLIGSPGRTFPVHPDLVMDARALACVYESLSKSRPGTASSSGGS--IARGAPVI	1529
CC	as nutritional supplements. They may be used to increase stem cell	Db	1003	ITTTIMNKPET-----AKPKDRATNSKATTPKPKOKTKAPKPTTKPKPT	1050
CC	proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon	QY	1530	VPELGKPRQSPLTYEDHGFAGHLPGRSPVTMRE--PTPLQEGSLSSS---KASQDKR	1584
CC	and/or nerve tissue growth or regeneration; immune suppression and/or	Db	1051	MPVRKPKTTP-----TPRKWTSMPELNPTSRIAEAMLQTTTRPNQTPNSK	1097
CC	stimulation; as anti-inflammatory agents; and in treatment of leukaemias.	QY	1585	LT-----STPREIAKSPHSTVPEHHPH-----PISPYEHLRLGVSGVDLYRS	1626
CC	AAU29510-AAU33304 represent the amino acid sequences of novel human	Db	1098	LVEVNPKSEDAAGAGETPHMLLR--PHVFMPEVTPDMVDYLPVNPQGIIN-----	1147
CC	secreted proteins of the invention	QY	1627	HIPLAPDPTSIPIRGIPLD 1644	
XX	Sequence 1415 AA;	Db	1148	--PMLSDETNICNGKFPD 1163	
Query Match 3.2%; Score 428; DB 4; Length 1415;					RESULT 68
Best Local Similarity 20.8%; Pred. No. 4.7e-15;					AAW78710
Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;					ID AAW78710 standard; protein; 2063 AA.
QY	476 NENYKSLVRSYRRRKQSQ-----QQQQQQQQQQQQQQQPMRPSQEEKDEKE	524	XX	AAW78710;	AC AAW78710;
Db	137 SQTIKSTTKRSPKPPNKKTKKVIIESEITEHVSVENQESSSSSSSSSTIWKIKS	196	XX	06-NOV-2001 (first entry)	DE Human protein SEQ ID NO 1372.
QY	525 KEKEAEKE-BEKPEVENDKEDLLKEK-----TDDT-SGEDNDEKEAVASGRKTANSQ	575	XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
Db	197 KNSAANRELQKLVKDNKNRKKTKPKFPVWDEAGSLDNGDFKVTPTDSTTQHNK	256	XX	tissue growth factor; immunomodulatory; cancer; leukaemia;	KW nervous system disorder; arthritis; inflammation.
QY	576 GRKGRITRSMANESEAITP-QQSABLASMELNESSRWTEEMETAKGLLEHGRNW	634	OS	Homo sapiens.	
Db	257 VTSPPKIT--TAKPINRPSLPPNSDTSKETSITVNETKTTTNTKQTSSTDGKEK	314	XX	WO200157190-A2.	
QY	635 SAIAIRMGSKTVSOCKNFYFNVKRQNLDEILQOHLKME---KERNARRKKKKAPAAAS	691	XX		
Db	315 TTSAKETOIEKTSADL-----APTSKVLAKPTPKAETTKGPAITTKPTTPK	367			
QY	692 EBAAPFPVVEDEMEASGVSGNEEMVEBAEALHASGNEVPVRCGSGPATVNNSSDTSIE	751			
Db	368 EPASITP-----KEPT-----PTTIKSAPTPKE	391			
QY	752 PSPHTEAAKDTGONGPKPATIGADGPPPGPTTPPRTSRAPITPEASEATGATPP--	809			
Db	392 PAPTITKSAPITPKKEPAPTT-----KEPAPTTKPEPAPTTTKEPAPTT-TKSAPITPK	445			
QY	810 PAPPSAPPPVVPVVEKEEKEEATAAPPVVEGEQPPAAELAVDTGKAEEVPKSECTEE	869			
Db	446 PAPTTPKKEPAPTPKE--PAPTTPKEPTTTPKEPAPTTKPEPAPTTKPEPAPTT-----	496			
QY	870 ABEGPAKGDAAEAATAGALKAEKKEGSGRATTAKSSGAPQSDSATTCSADEVDEA	929			
Db	497 APKKEPAPTTKPEPAPTTKPEPAPTTTKE-----PSPTTPKEPAPTTTKSAPITTKPE	548			
QY	930 EGGDKNRLISPRPSLLTPTGDPANASPKQLDLKQKRAAAIPIQVTKVHEP-----	985			
Db	549 -----APTITKSAPTTKPEP-----SPTTKEPAPTTTP	576			

QY 1932 LME-----FVLLKEAPRVARPERPRADTGHAFKAPKAPASGKSEPRP--- 1983
D 1279 ILKAIQAPSNLTWNSFNATPQTHKLDVSVVNSGK-QNSGATKASPSNRRSPGSS 1337
QY 1984 -LVPPVSGHATARTAKNLA-----PHH--ASDPDPA-PPASASDPHREKT 2026
D 1338 RKTTPSPGRON-SKAPKLTASOTNALQONVELPRNVLSPTPLANPPVGSFPNNS-- 1394
QY 2027 QSKPFSIQELSLGSHGSSVPEGVPSVSSPSLTHDKGLPKHLELDKSHLEGEL 2086
D 1395 -----GLNPQNSTVSAVAGGVVDENKESLNVPODSCQNSQRKEQVNIEL 1441
QY 2087 RPKQPGVKLGGEAHLPLRLPESQPSPLLOTPAGVKHQHVVVTLAQHISEVITQD 2146
D 1442 KAVPAQEVKM-----VVPEDSKQDQSDPNKL-----PSVEENKLVSPAM----- 1484
QY 2147 YTRHHFQQLSAPLPAPLYFPFGASCVDLRLRP-SDLYLPPDPHGAPARGSPHSGGR 2205
D 1485 --REAPTSLSQLL-----DNSGAPNVTIKPGLTDLEVTTP----- 1518
QY 2206 SPEPKNTSVLGGEGDIEPVSPEGTEPHGSRSAVYPLLYRDGEQTEPSRMGSKSPGNT 2265
D 1519 -----VSGED-----LKKASVITPL-----QDLSSSKPEPSNS 1546
QY 2266 SOPPAFFSKLTESNANVKSQKQEIINKLNTNHRNEPEYNIQPGTEIFNMPAITGTGLM 2325
D 1547 LNLPL-----HSNELCSSLVHPELSEVSS-----NVAPSIIPVMSRP-VSSSSIS 1589
QY 2326 TYRQAOVQEHASTNMGLEAIRKALMGKYDOWEESPPLSANAF---NPLNAGASLPDAMP 2382
D 1590 T-----PLPPNOITVFTVSNPITTSANTSALPL 1617
QY 2383 ITAADGRSDHTLTPGGGKAKVS-----GRP-----SSRKAK 2415
D 1618 THLOSALMTVVTNFMNAGSKVMVSEQSAQSNARQFPTIPFINSSIIQWKGSPBT 1677
QY 2416 SPAPGLA--SGDRPPSVSV---HSEGDGN-RRTPLTNRVWEDRFSAGSTPPFY----N 2465
D 1678 IDAAPTNTNSGLMPPSVAVVGLPHIQNIKFSAPVP-----PNALSSPAPNIQTGR 1730
QY 2466 PLIMRLQAGVMAAPPPGIPAGSGPLAGPHAWDEPKPL 2505
D 1731 PLVLSTRATFVLQLEP---PCTSSPVV-FSHPPVQVQVKEL 1766

RESULT 69

AA040064
ID AA040064 standard; protein; 2063 AA.
AC AA040064;
XT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 3209.
XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
OS Homo sapiens.
XX WO200153312-A1.
PN 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US034263.
PF 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AAI59220.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
PS Example 5; SEQ ID NO 3209; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
encoded polypeptides (AAM38642-AAM42213) with neurotropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemia and
C.N.S disorders. Note: The sequence data for this patent did not form
part of the printed specification
XX Sequence 2063 AA;
Query Match 3.2%; Score 428; DB 4; Length 2063;
Best Local Similarity 19.3%; Pred. No. 7.7e-15;
Matches 413; Conservative 272; Mismatches 849; Indels 606; Gaps 99;
QY 591 NSEAITPQGAELASAMELNESSRWTEEMETAKGLLEHGRNWSAIAIWMGSKTVSOCK 650
D 8 NLEDIYTSICSTWEDSEDFDGL--EDDTKSDSILEDSTIFVAFKGNIDDKD----- 60
QY 651 NFYFNKKRQNLDEYL-----QQHKLKVEKER--NARRKKKKAPAAAEAEAFPPV 699
D 61 ---PKWK---LDAILKNVPNLLHMESSKLVQKVEPNVSVRTENI PREAAERLRILAQ 113
QY 700 VEDEMEASGVSGNEEEMVEAEALHASGNEVPRGECGSPATVNNSSSTESIPSHTEAA 759
D 114 SNNQQLRDLGILSVQIE--GEGAINLALQNRSDQVRMNGPMGAGNSVRMEA--GFPMSGP 171
QY 760 KDTGONGPK---PP-----ATLGADGPP--GPPTPPRRTSRAPIER-----TPASE 801
D 172 GIIRMNNTATWIPPGNVSSMWAPGNPLQPTPRPASQSDAMDPLLSGLHTQQOSH 231
QY 802 ATGAPTPPPAPPSP-----SAPPVVPKKEKEETAAAPVVEEGEQKPPAAELAVDTG 856
D 232 PSGSLAPPHHPWQPVSVNVRNMPANFQLOQQQQQQQQQQQQQQQQQQQQQQQQARPP 291
QY 857 KAEEPVKSECTEEAEAGPAKGDAAEAEATAEGALKAEKKEGSGRATTA---KSSGAPQ 913
D 292 QOHQQQQQGGIRPOFTAPTQVPPGWNQLPSGALQPPPAQSGSLGTMTANQGWKKAPLP- 350
QY 914 DSDSSATCSADEVDEAEGGDKNRLISPRPSLLTPTGDPANASPOKPLDLKQLKORAAAI 973
D 351 -----GPMQQQLQARPSLAT----- 365
QY 974 PPIQVTKVHEPPREDAATKPAFPAPPPQNLPQESDAPQ--PGSSPRKGRSPAPPADK 1032

[illegible]

1279	ILKAIQAPSNLTMPNSFATPOTHKLDLGVVNSGK-QSNSGATKCAASPSNRRSRSPGSS	1333
1984	-LVPPVSGHATIAARTPAKNLA-----PHH--ASDPDPA-PPASASDPHREKT	2026
1338	RKTTSPGRQN-SKAPKLTLASQTNAAALQNVELPRNVLSVPTPLANPPVPGSFPNNS--	1394
2027	QSKPFSIQEULELSLGYHGSYSYSGEVPSPVSPSLTHDKGLPKHLBELDKSHLEGL	2086
1395	-----GLNPQNSTVSAAVAGGVVEDNKESLNVPODSCQNSQSRKEQVNIEL	1441
2087	RPKQCPVVKLGGEAAHLPHLRPLPESQSSPLQTAGVKGHORVVTLLAQHISEVITQD	2146
1442	KAVPAQEVKM-----VVPDQSKKQCPQDPNKL-----PSVEENKMLVSPAM	1484
2147	YTRHHQQOOLSAPLPAPLAPFPGASCPLDLRRPP-SDLYLPDPDHGAPARGSPHSGGKR	2205
1485	--REAPTSLSQLL-----DNSGAPNVTIKPPGLTDLVTPP-----	1518
2206	SPEPNKTSVLGGEDGIEPVSPPEGMTPEGHSRAVYPLYRDGQTEPSRMGSKSPGNT	2265
1519	-----VVSGED-----LKKASVIPTL-----QDLSSSKSPNS	1546
2266	SQPAPEFSKLTESNAWKSKQEIINKLNTNREPEYNI SQPGETEIPNMPAITCTGLM	2325
1547	LNL P-----HSLNCLSLVHPELSEVSS-----NVAPSI PPVMSRP-VSSSSIS	1589
2326	TYRSQAVQEHASTNMGLEAIIRKALMGKYDQWEESPPLSANAF---NPLNASASLPAAMP	2382
1590	T-----PLPPNQITVFVTSNPITTSANTSALP	1617
2383	ITAADGRSDHTLTPSGGGKAKVS-----GRP-----SSRKAK	2415
1618	THLQSALMSTVVTMPNAGSKVMVSEGSAAQSNARQFITPVPFINSSIIQVMKGQSPST	1677
2416	SPAPGLA--SCDRPPSVSV---HSEGD CN-RRTPLTNRWEDRPSAGSTPPY---N	2465
1678	IPAAPLTNLSGLMPPSVAVVGPLHI PQNIKFSAPVP-----PNALSSPAPNIQTGR	1730
2466	PLIMRLQAGVMASPPPPGGLPAGSGPLAGPHHAWDEBP KPL	2505
1731	PLVLSSRAFPVQLPSP---PCTSSPVV-PSHPVPVQVQVKEL	1766
RESULT 70		
ABG17147		
ID	ABG17147 standard; protein; 2063 AA.	
XX		
AC	ABG17147;	
XX		
DT	18-FEB-2002 (first entry)	
XX		
DE	Novel human diagnostic protein #17138.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
OS	Homo sapiens.	
XX		
XX		
PN	WO200175067-A2.	
PD	11-OCT-2001.	
XX		
XX	30-MAR-2001; 2001WO-US008631.	
PF		
XX		
PR	31-MAR-2000; 2000US-00540217.	
PR	23-AUG-2000; 2000US-00649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Drmanac RT, Liu C, Tang YT;	
XX		
XX	WPI; 2001-639362/73.	
DR	N-PSDB; AAS81334.	
DR		

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX

PS Claim 20; SEQ ID NO 47506; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 2063 AA;

Query Match 3.2%; Score 428; DB 4; Length 2063;
Best Local Similarity 18.9%; Pred. No. 7.7e-15;
Matches 408; Conservative 269; Mismatches 837; Indels 646; Gaps 98;

QY 591 NNEEATPOOSAEALAMELNESSRWTEEMETAKGLLEHGRNWSAIAEMVSGKTVSQCK 650
DB 8 NLEDIYTSICSTMEDSEDFDGL--EDDDTKSDSILEDSTIFVAFKGNIDDKD----- 60
QY 651 NFVFNKQNLDEIL-----QOHKLKWEKER--NARBKKKKAPAAASEEAFFPV 699
DB 61 ----FKWK-----LDAILKVNPLHMSKLLKQKVEPWNVSRVTNIPREAAERLILAQ 113
QY 700 VEDDEEASVGSNEEMVEAEALHASGNEVPRGECGPATVNNSSDSTESIPSHTEAA 759
DB 114 SNNQQLRDLGILSVQIE-GEGAINLALAQNRSDQVRMNGPMGAGNSVRMEA-GFPWASGP 171
QY 760 KDTGQNGPK-----ATLGADGPPP--GPPTPPRTGRAPTEP-----TPASE 801
DB 172 GIRMNNPATVMPFGGVNVSNNMAGPNPELPQRTPRPASQSDAMDPLLSGLHIQQQSH 231
QY 802 ATGAPTPPPAPPSP-----SAPPVVPKEKEETAAAPVVEGEQKPPAAEEELAVDTG 856
DB 232 PGSGLAPPHHPQVSVNQMNPANFPQIQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 291
QY 857 KAEPPVKSECTEAEAGPKAGDAEAAEATAGALKAEKKGSGRATTA---KSSGAPQ 913
DB 292 QQHQQQQQPGIRPQFTAPTQVPPVPGWNLPSGALQPPPAQGSLSGTMTANQGWKXAPLP- 350
QY 914 DSDSSATCSADEVEDEAGDKNRLSPRSLTPTGDDPRANASPOKPLDLKOLKORAAAI 973
DB 351 -----GPMQQQLQARPSLAT----- 365
QY 974 PPIQVTKVHEPPREDAAPTKPAPPAPPONLQPSDAPQ--PGSSPGKSSRAPPADK 1032
DB 366 --VQT-----PSHPPPPYFGGQQQASQANTNPMQNSPCQFTAP----- 402
QY 1033 EAFAAEAOQLPGDPPCWTSGLPFPVPPREVIVKASHPADPSAFSY-APPGHPLPLGLHDT 1091
DB 403 -----QMSLQGGP-----SRVPTLQOQPHLTNKSPP-ASSPSSFQGGSPASSPT---VNQT 449
QY 1092 ARPVLPRPTTISNP-----PPLISSAKHPSVLEIRQIGAISQMSQVQLHVPVSEHAKAPV 1145

DB 450 QQOMGPRPQ--NNPLPQGGQQQVSPSGRPNMV---QQGNVPPNFMVMOQQPQNO----- 499
QY 1146 GPVTMGLPLPMDPKKLAP--FSGVKOBQLSPROQAQPPESLG-VPTAQEASVLRGTALGVS 1203
DB 500 GPQSLHPLGLGMPKRLPPGFS-----QANPNFMQGGVP-----STTATT 540
QY 1204 PGGSITKGTPSRVPSDSAITYRGSIITHTPADVLYKGTITRIIGEDSPSLDRDREGDSL 1263
DB 541 PONS---GAP--OLQANQNVHAGGAGGAPPQNM-----QVSHGPPNMM 580
QY 1264 PKGHVIEGKKGHVLSEYSGMSVTQCKEDGRSSGPPHETAAPKRRTYDMMEGRVGRAIS 1323
DB 581 QPSLMGIHGNMNNQAGTSGVQVNLNMQGQPP----- 618
QY 1324 SASIEGLMRAITPRRHHPHLKEQHHTIRGSTQIGIPRSYVEAQEDYLREAKLL--KR 1380
DB 619 -SQLMGHMQIIVP-----SQGMVQQ---QGLTNQNPMLLSRAQ---LMPQGMVNPSP 667
QY 1381 EGTTPPPP---PPSRDLTE-----AYKTOALGP-----LKLKPAHEGLVATVKEAGRSI 1425
DB 668 QNLGSPQBMTPPKQMLSQGGPQWMAHPNQMMPGQGVLLQONPMIEQIMTNMQGNKQQ 727
QY 1426 HEIPREE--LRHTPELPLAPRLKEGSITQGTPLKYDTGASTTGSKKHDVRSLSIGPRT 1483
DB 728 FNTQNSQNVMPGPAQIMRGPTNMQNMVQFTQMSQGMQLPQGGPVNNSPSQVMGIQGGV 787
QY 1484 FPPVHPLDMADARALERACYEESLSRPGTASSGGSIARGAPVI-----VPELGRP 1536
DB 788 LRPPGSPHMAQOH-----GDPATTANNVDSLSQMMPDVSIQOTNMVPHVQA 835
QY 1537 RQ-----SPLTYEDHG-----APFAG-----HLPRGSPVTMREP--TPRLQEGSL 1574
DB 836 MGNNSASGNSHPSGCHGMSFNAPFSGAPNGNQMSCQNGPFPVNVKVTLSPLLVNLLQSDI 895
QY 1575 SSKKASQDKJTS-----PREIAKSPHS--TVPEHHPHPTISPYEH--LLRGVSGVD 1622
DB 896 SAGHFGVNNKQNTNANKPKKKKPPKKKNSQDQDLNTPDTRPAGLEADQDPLPGEQGTS 955
QY 1623 LYRSHIPLAFDPTSTPRGI---PLDAAAAYLP---BHLAPNTYHYLP--PYLIR--- 1671
DB 956 LNSG--PKLPFSNPPPGYPSQVPEQRPLQOMPQPMQHVAPPPOPPQQQPPQPPQQQ 1014
QY 1672 ---GYPTAALENRQTTIINDYITSQQMHNTATAMAQADMLRG--LSPRESSLALNAA 1726
DB 1015 PPPSPQSQ 1074
QY 1727 GPRGIIDLSQVPHLPVLPPTP-----GTPATAMDRLAYLPTAQ--PPSS----- 1770
DB 1075 SVPMVSLQG---PASVPPSPDKQRMPPMVPNTPLGNSNRKMWVQESQNPSPSSPLAEM 1130
QY 1771 ----RHSSSPLSPCGGPHLTKPTTTSSSERDRDRDREREKSIILSTTTTVEHAP 1825
DB 1131 SLPEASGSEAFSVPGGPNMFSHV-----VLPQNQLMTGTP 1166
QY 1826 IWRPQTEOSSSGSSGGSSGSSSRPASHSHAHQSP--ISPRTOALQORPSVLHNTGMK 1884
DB 1167 --KPGSPSLATQCATPQPPVNSLPSHGH--HFPNVAAPTQTS---RPKTPNRSRSP 1218
QY 1885 GIITAVEBSKTVLRSTSTS--SPVRPAATFPPTHCPGLGTLGDVYPTLMEPVLLPKAP 1943
DB 1219 FYYPQTPNNRPPEPSEISLSPER-----LNASIAGLFP-----P 1254
QY 1944 RVARPERADTGHAFKAPARSGLEBPASSPSKSGSEPRPLVPPVSGHATIAARTPAKNA 2003
DB 1255 QINPLPPRPNLNRGF-----DQQLNPTTLKAIQAPSNLTMAPSNFAT----- 1299
QY 2004 PHASPDPPAPPASADPHREKTSKPFISIOELELRSLGYHSGSSYPSGEGVPSVSPSS 2063
DB 1300 -----PQTHKLDVVVN-----SGKQNSGATKRASPSNS-- 1329
QY 2064 LTHDKGLPKHLELDKSHLEGELRPQKQPVKLGGEEAHLPHLRPLPESQSSSPLLTQA 2123

Db 541 PGNS---GAP--OLQANQNVQHAGGQAGPQNM-----QVSHGPPNMV 580
QY 1264 PKGHVIEGKKGHLVSEVGMSVTQCSKEDGRSSGPPHETAAPKRTYDMMEGRVGRATS 1323
Db 581 QPSLMGIHGNMNNQAGTSGVPOVNLNWCQGOQGGP-----618
QY 1324 SASIEGLMGRAIPERHSHLKEQHIIHRSITQIGIPRSYVEAQEDYLREAKLL---KR 1380
Db 619 -SOLMGHGOIIVP---SQOMVQV---QGLNPNQPMILSRAQ---LMPQGMVNPVS 667
QY 1381 EGTTPPP---PSSRLTE-----AYKTCALGP-----LKLPAHEGLVATVKEAGRSI 1425
Db 668 QNLGSPQRTWTPKQMLSQOGPQMAMPHNQMGPOGVLLQQNPMTIEQIMTNQMQNKKOQ 727
QY 1426 HEIPREE--LRHPELPLAPRLKEGSIQTGLKYDTGASTGSKKHVRSLSIGSPGRT 1483
Db 728 FNTQNSNVMPGAQIMRGFTPNMGVMVQFTQMSGQMLPQCGPVNNSPSQVMGLQGV 787
QY 1484 FPPVHPLDVNADARALACYEESLKSFRGTASSGGSIGARGAPIV-----VPELGKP 1536
Db 788 LRPPGSPHMAQQH-----GDPATTANNDVLSQMPDVSIQOTNNVPPHVQA 835
QY 1537 RQ-----SPLTYEDHG---APPAG-----HLPRGSPVTMRPE-TPRLOEGSL 1574
Db 836 MQNSASGHNHFGHGMSFNAPFSGAPNGNQMSCQNPFGPFVNKDVTLTSPLLVNLQSDI 895
QY 1575 SSKASQDRKLTST-----PREIAKSPHS-TVPEHHHPHPISPYEH-LLRGVSGVD 1622
Db 896 SAGHGVNNKONTNANKKKPKPKKNSQODLNTPTTRFAGLEADQPFLPGEQGIS 955
QY 1623 LYRSHITLAFDPTSPRGI---PLDAAAAYLP---RHLAPNPTVPHLYP-PYLIR--- 1671
Db 956 LDNSG-FKLPEFENRPPGVSQVEQRLQOMPQLMQHVADPPQPPQPPQPPQPPQ 1014
QY 1672 ----GYPTDAALENROTIIINDITYTSQMEHNTATAMAQADMIRG-LSPRESLALNYAA 1726
Db 1015 PPPPSQFQQQQQQQQQQMMMLMQDDPKSVLQVNSQNVHPHPRGPNLNDPDRMPMQSG 1074
QY 1727 GPRGIIDLSQVHLPLVLPPT-----GTPATAMDLAVLPTAPO-PFSS----- 1770
Db 1075 SVPMVSLQ-----PASVPPSDKQRMMPVNTPLGNSRKVMYQSPQNPSSPLAENVA 1130
QY 1771 ----RHSSSPSLPGGTHLTKTPTTSSERERDRDRDREREKSLTSTTTVEHAP 1825
Db 1131 SLPEASGSEAPSPVCGPNNMPHV-----VLPQNLMMTGP 1166
QY 1826 IWRPGTEQSSGSGSGSGGSSRSPASHSHAHQSP-ISPRDQALQORPSVHLNTGK 1884
Db 1167 --KPGSPLSATQAGTQPPQPPVNSLPSHGH---HFPNVAAPTQTS---RPKTPNRASPR 1218
QY 1885 GIITAVEPSKPTVLRSTST-SFVRPAATFPFATHCPLGGLDGVVPTLMPEVLLPKEAP 1943
Db 1219 PYPQTPNNRPSTPSEISLSPER-----LNASTAGLFP-----P 1254
QY 1944 RVARPERPADTGHAFAPKAPRAGLEPASPSKSEPRPLVPVPSGHATARTPAKNIA 2003
Db 1255 QINIPLPRLNLRGP-----DQGLNPTTLKAIGQAPSNLTWNPSNFAT----- 1299
QY 2004 PHASDPDPAPASADPHREKTQSKFISIOELRLSLGYHSSYSPEGVPSVPSVSSPS 2063
Db 1300 -----POTHKLDVVVN-----SKQNSGATKCAASPSNS-- 1329
QY 2064 LTHDKGLPKHLEBLSKHELRPKQPGVKLGGEAAHPLHLRLPESQPSSSPLQTA 2123
Db 1330 -----RRSSPGSSR-----KTPSPGQNSKAPKL--- 1354
QY 2124 PGVKGHRVVTVAQHISEVITDYTHHPQQLSAPLAPLYFPGASCVLDRPPSOL 2183
Db 1355 -----TLASQTNAAALQ-----NVELPRNVL 1375
QY 2184 YLPPPDHGAARGS-PHSGGGRKSPBNKTSVLGGG---EDGIEPVSP- 2228
Db 1376 VSFTPLANPVPGSPFNNSG--LNFQNSTVSVAAGVVVEDKESLNVPDQSDCQNSQR 1433

QY 2229 -----EGMTEPGHSRSAYVP--LLYRDGEOTEPSRMGSKSPGNTSQPPAFPSKLTESNS 2280
Db 1434 KEQVNIELKAVPAQEVKVMVPEDQSKDQSPDNKLP- 1477
QY 2281 AMVSKKQKQINKKLN--THNRNEPYNISQPG-TEI-FNMPAITGTGLM-----TYRSQ 2330
Db 1478 NLVSPAMREAPTSLQLLDNSGAPNVTIKPPGLTDLVTPPVVSGEDLKKASVITLQDL 1537
QY 2331 AVOEHASTNMGL-----EAIIRKALMKYDQWEE- 2365
Db 1538 SSKSPSLSNLPHNELCSSLVHPSELSEVSNVAPSPVPMVRPVSSSSISTPLPPNOI 1597
QY 2366 NAF--NPLNASASLPAAMPITAAGRSDHTLTSFGGCKAKVS-----GRP--- 2409
Db 1598 TVFVTSNPTTSANTSAALPHTLQALMSVTVTMFMNAGSKVMVSGQSAQSNARPOFIT 1657
QY 2410 -----SSRKAKSPAPGLA--SGDRPPSVSV---HSEGDEN-RRTPLTNRV 2449
Db 1658 PVFINSSSIQVMKSGSPSTIPAAPLTTNSGLMPPSVAVVGPLHIPQNIKFSSAPVP--- 1714
QY 2450 WEDRESSAGSTPPY-----NPLIMELQAGVMASPPPPGLPAGSGPLAGPHHAWDEPKPL 2505
Db 1715 ----FNALSSSPAPNIQTRPLVLSSRATPVQLPSP---PCTSSPVV-PSHPVQVVKEL 1766
RESULT 72
ID ABB98406
XX ABB98406 standard; protein; 2545 AA.
AC ABB98406;
XX
DT 21-OCT-2002 (first entry)
XX
DE Human NOV5, MAST205-like protein.
XX
KW Human; NOV5; cytostatic; Cardiant; Antiinflammatory; Immunosuppressive;
KW Antiallergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;
KW Antiasthmatic; Nephrotropic; Hepatotropic; Neuroprotective; Nootropic;
KW Antibacterial; Virucide; Antiparasitic; Relaxant; Anticonvulsant;
KW Gene Therapy; NOV; cancer; heart disease; inflammation;
KW autoimmune disorder; allergy; blood disorder; AIDS; diabetes; obesity;
KW asthma; IGA nephropathy; cirrhosis; arthritis; Alzheimer's disease;
KW infection; stroke; muscular dystrophy; epilepsy; wasting disorder;
KW MAST205-like protein.
OS Homo sapiens.
PN WO200255704-A2.
XX
PD 18-JUL-2002.
XX
PF 09-JAN-2002; 2002WO-US000554.
XX
PR 09-JAN-2001; 2001US-0260417P.
PR 10-JAN-2001; 2001US-0260831P.
PR 28-FEB-2001; 2001US-0272338P.
PR 09-MAR-2001; 2001US-0274876P.
PR 18-APR-2001; 2001US-0284704P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Padigar M, Li L, Zerhusen BD, Casman SJ, Shenoy S, Spytek KA;
PI Zhong M, Gangolli EA, Burgess CE, Patturajan M, Vernet CAM;
PI Taylor S, Tchernev VT, Miller CE, Guo X, Boldog FL, Grosse WM;
PI Alsobrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;
PI MacDougall J, Malyankar U, Millet I, Peyman J, Smithson G;
PI Gunther E, Stone DJ;
XX
DR WPI; 2002-590674/63.
XX N-PSDB; AEN85383.
XX
PT NOVX polypeptides and encoding polynucleotides, useful for preventing or

Db 2222 -----EARGKPG-PKPP-----TEAD-----KPNGMKRSP----- 2247
QY 2308 QPCTEIPNPAITGTGLMYRQAVOEH-----ASTNMGLEAIRKALMGYDQWESPPL 2363
Db 2248 -----SATGSSFRSTALPEKSLSCSSFPETAGVREASASSD----- 2287
QY 2364 SANAFNPLNASLPAAMPITADGRSDHTLTSPGGGK---AKVGRPSSR----- 2412
Db 2288 -----TSSAKAAGCMLELPAPNRDHRKAQPAGEGRTHMTKSDSLPSFRVSTLPLES 2339
QY 2413 -----KAKSPAPGLASGRDRPSVSVHSGDCNRTPL 2445
Db 2340 HHPDPNMGASHRDRLSVTATVGTGKOPAPA-----QPPARKONVGRDVTKPSPA 2394
QY 2446 TNRVWEDRPSS 2456
Db 2395 PN---TDRPIS 2402

RESULT 74
AAU03503
ID AAU03503 standard; protein; 2523 AA.
AC AAU03503;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human protein kinase #3.
XX
KW Human; protein kinase; PK; STK; cancer; cardiovascular disease;
KW metabolic disorder; immune related disease; neurological disorder;
KW neurodegenerative disorder; inflammatory disorder; infectious disease;
KW reproductive disorder.
XX
OS Homo sapiens.
XX
PN WO200138503-A2.

XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WO-US032085.
XX
XX 24-NOV-1999; 99US-0167482P.
XX
XX (SUGEN) SUGEN INC.
XX
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI Flanagan P, Clary D;
XX
DR WPI; 2001-343950/36.
DR N-PSDB; AAS06703.
XX
XX Nucleic acids encoding human kinase polypeptides, useful for preventing
PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and
PT neuronal-associated diseases, and microbial infections.
XX
PS Claim 7; Fig 2; 433pp; English.

XX
XX AAU03501-AAU03557 represent novel human protein kinases #1-57. The novel
CC protein kinases have been identified as members of the tyrosine or
CC serine/threonine kinase (PK and STK) families. The polynucleotides
CC encoding protein kinases and the polypeptides may be used in the
CC prevention, diagnosis and treatment of diseases associated with
CC inappropriate kinase expression. For example, they may be used to treat
CC cancers (especially cancers of haematopoietic origin), cardiovascular
CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
CC immune related diseases (e.g. rheumatoid arthritis), neurological
CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).
CC Additionally, polynucleotides encoding protein kinases may be used for
CC gene therapy and as DNA probes in diagnostic assays. The protein kinase
CC polypeptides may be used as antigens in the production of antibodies

CC against the protein kinases and in assays to identify modulators of
CC protein kinase expression and activity
XX
SQ Sequence 2523 AA;
Query Match 3.2%; Score 424.5; DB 4; Length 2523;
Best Local Similarity 19.8%; Pred. No. 1.6e-14;
Matches 549; Conservative 308; Mismatches 979; Indels 935; Gaps 139;
QY 38 GLELVQH-----SDYASHLSPGSI-----LQORRRPRLSEFPQGNERSQELHLRPE 87
Db 309 GVLSTFHQIIBLARDCLDKSHQGLITSRYFLEQHKDKLL-----QEAHRDSE 358
QY 88 SHSYLPGLKSMETIESKRPRLELLPDLPLSPPLATGQAGSEDLTKDRSLTKCLRP 147
Db 359 S-----GELAFIKQLVRKILV---IARPARLLEC-----LB- 387
QY 148 VSPPPPHTDPE-----LELVPRLSKEELIQMDRDVREITMVEQIISKLKKQOOLEE 202
Db 388 -----FDPEFYVLLAEAGHAKGQGIKT--DIPRYI-----ISQLGLNKDPLEE 431
QY 203 EA-----AKPEPEKPVSPPIESKHSLSVQIYDENRKKABAAHRILEGLPQVE 253
Db 432 MAHLGNYSGTAEPTETDESVSNSASLKL-----RKPRE----- 467
QY 254 LPLYNQPSDTRYHENIKI--NOAMRKKLILYFKERNHARKOWKOKFCORY-----DQLM 306
Db 468 -----SD-----FETIKLSNGAYG---AVFVHKESRQRFAMKKINKQNLILNQIQ 513
QY 307 EA-LBKVKVERIENPRRAKESKVREYKQPEIRKQELQERMOSRVG-----QRGS 359
Db 514 QAFVERDILTFAENP-----FVVSVCSPETRRHLCMVMEYVEGGDCATLMKMN 562
QY 360 G-LSMSAARSEHVESEIIDGLSEQENLEKQORLAVIPMLYDADQOQIKFINMN-----G 414
Db 563 GPLPVDMAR--YPAETVLALEYLHNYGIVHRDLK--PDNLLVTSMGHIKLTDFGLSKVG 618
QY 415 LMADPMKVKY-----DROVMNMWSEQEKETFREKFMQHPKNGFLIASFLERKTV 463
Db 619 LMSMTNLVEGHIEKDAREFLDKQVCTPEYIAPEVILRQGVKPVDMWAMGIIILYFVL 678
QY 464 AECVLYYLT-----KNE-----NYSLV-----RRSVRRGKSQOQOQ 498
Db 679 G-CVPFFGDTPEELFGQVISDEINWPEKDEAPPPDAQDLITILLRQNLERLGTGAVEV 737
QY 499 QOQOQOQOQOQOQPMPSRQOEKDEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 557
Db 738 KQHRFRSLDWNLLRQKAEFIQLESEDDTSTYFTRSEKYHME---TEEEDDTNDEDF 794
QY 558 NDEKEAVAS---KGRKTANSQORRRGRITRSMANSEAEAITPOQSAELASME--LNESS 613
Db 795 NVEIRQFSSCSHRFSKVPSS---IDRITONSABE--KEDSVDTKSTTLPLSTETLSWS 848
QY 614 RTE--EEMETAKGLLEHGRNWSAIARMVGSKTVSQCKNFYNYKRONLBILQOHLK 672
Db 849 EYSEMQQLSTNSNSDTSNR-----HKLSSGL---LPLKLS 882
QY 673 MEKENARRKKKAPAAASAEAPPPVVEDEMEASGVSGNEEMVEAEALHAGSNEV- 731
Db 893 TEGEQD---EAASCPGDPHEEPGKPALPPE-----CAQEEPEVTTTASTISSSTLSVG 933
QY 732 -----PRGCGSPATVNNSSDTSIPSPH-----TEAAKDTGONGPPTATL 773
Db 934 SFSEHLQDQINGRSECDV---STDNSSKPSSEFASHMARQLESTKKSIG---KVTKSL 987
QY 774 GADG-----PPGPPTPRRTSRAPIEPTPASEATGATPPPPAPSPSPAPPPV 821
Db 988 SASALSILMPGDMFAVSLGSPMSPHLSL-----SDPSSSRDSSPSRDSASAASHPOPI 1042
QY 822 VPKEKEEE--TAAAPVVEEGE-----EOKPPAAE----- 849
Db 1043 VHSCKNGYGTIRAIRVYVGDSDIYTVHHVWVEGSPACQAGLKAGDLITHINGEPV 1102

FT Misc-difference 121 /note= "Lys encoded by GAC"
 FT Misc-difference 122 /note= "Asp encoded by AAG"
 FT Domain 448..543 /note= "SH3 domain"
 FT Misc-difference 532 /note= "Phe encoded by CCT"
 FT Domain 587..684 /note= "PDZ domain"
 XX WO200078921-A2.
 XX 28-DEC-2000.
 XX 23-JUN-2000; 2000WO-US017322.
 XX 24-JUN-1999; 99US-0140715P.
 XX (UJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 XX (MASS-) MASSACHUSETTS GEN HOSPITAL.
 XX Worley P, Tu JC, Xiao B, Sheng M;
 XX WPI; 2001-102712/11.
 XX N-PSDB; AAF24877.
 XX New Shank polypeptide for identifying Shank protein modulator compounds
 XX used to treat Alzheimer's disease and stroke, comprises an ankyrin
 XX domain, SH3 domain, PDZ domain, proline-rich domain, and SAM domain.
 XX Claim 4; Fig 2A; 118pp; English.
 XX The present sequence represents a rat Shankla polypeptide. The
 XX specification also describes Shank2 and Shank3a. The Shank polypeptides
 XX have an ankyrin domain, SH3 domain, PDZ domain, proline-rich domain, and
 XX SAM domain. Shank polypeptides play a significant role in the post
 XX synaptic density cytoskeleton. Compounds that modulate Shank proteins are
 XX useful for treating a disorder associated with glutamate receptors such
 XX as epilepsy, glutamate toxicity, disorders of memory, disorders of
 XX learning, stroke, schizophrenia, Alzheimer's disease, tissue degeneration
 XX and disorders of brain development. A disorder associated with Shank
 XX protein activity such as cardiac disorder, disorder of musculature,
 XX vasculature disorder, neurological disorder, psychiatric disorder, renal
 XX disorder, uterine disorder or a disorder of bronchial tissue is also
 XX treated by a compound that modulates Shank protein activity
 XX Sequence 2087 AA;
 Query Match 3.2%; Score 422; DB 4; Length 2087;
 Best Local Similarity 20.8%; Pred. No. 1.7e-14;
 Matches 378; Conservative 141; Mismatches 619; Indels 682; Gaps 83;
 QY 658 KRQNLDEILQHKLMKEKERNARRKKKAPAAASEEAPPPVVEDEME---ASGVSGNE 714
 DB 686 RHFDME-----AVHKASQQAQLPPALSLRSKMSWSELEMEYEQQAAPVSMK 738
 QY 715 EEMVEEA-----EALHAGSNEVPRGECGPATVNSSD-----TESIPSPHTEAAK 760
 DB 739 KRTVYQMALNKLDLILAAQQTISASESPGGLASLGKHKRPGKFATESFDPHRS--- 796
 QY 761 DTGQNGKPKPATLGADGPPGPPPTPRRTS-----RAPIETPA---SEATGATPPPA 811
 DB 797 ---QPSYDRPSFL-----PPGGLMLRQKSIAGAAEDDRPYLAPPAMKFSRSLVSGSEDI 848
 QY 812 PPSGAPPVPVPEKEEBEETAAAPVVEEGEEQKPPAAABELAVDTCKAEVPKSECTEAE 871
 DB 849 PPTPTSP-----EPY-----STPFA----- 866
 QY 872 EGPAGKDAEAAATAEGALKAEKGGSGRATTA-----KSSGAPQDSSSATCSAD 924
 DB 867 -----PSSSGLRTPRGGFNPPSSGGPLPASPSSPSFDGP 901

QY 925 EVDEAEGGDKNRLSPRPSLLTPTGPRANASPOKPLDLKQLKQRAAALPIQVTKVHEP 984
 DB 902 SPPTDTGGGKREKSL-----YHSAALPPAH-----HHP 928
 QY 985 P---REDAAPTTP-----APPAPPPQNLQPEDSDAPQOQSSPRGKR 1024
 DB 929 PHHHHHAPPQPHHHHAPPHPPEMETGGSPDDPPRLALGQPSLRGWRGGPSPSTSG 988
 QY 1025 SPAPPADKEAFAAEAQKLPDPCWTSGLFPVPPREVIKASPHAPDPSAFYAPP----- 1080
 DB 989 APSGSHSSSGSSG---PTQAPALRY---FOLPPR-AASAAAYVPASGRGRKGLVKQ 1041
 QY 1081 ----GHP-----LPLGLHDTARVLPRLPPTISNPPPLISSAKH-----PSVLER-----QIG 1123
 DB 1042 TKVEGEPOKGSIPSASSPTIS-PALPK-----SEPPAGPSEKNSIPITIIIIKAPSTNSG 1096
 QY 1124 AISQGMVOLHVPISEH-----AKAPYGVVTMGLPLMDPKKLAPFGVKQE 1170
 DB 1097 RSSQGSSTEAEPTQPDGAGGGSSPSAPATSPVPPSPSPVPTPASPSGATLDTSQF 1156
 QY 1171 QLSPRGQAGPPESLGVPTAQAEASVLRTALGSPGSGITKGIPTSTRVSDSAITVRSIT 1230
 DB 1157 GAALVGAARREGQWNEARRRSTLFLSTDAGDEGDSGLG-PGG--PPGRLRHSKSID 1213
 QY 1231 HGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEYEGKGVLSYEGGMSVTQCS 1290
 DB 1214 EG-----MFSAEPLRLLESG-----GSSGGYGAYAAG-----S 1241
 QY 1291 KEDGRSSGPPHETAAPKRTYDMGRRVGRATISSASISGLM-----GRAI 1335
 DB 1242 RAYGGSGSSSAFTSFLPPR--PLVHPLTKALDPASPLGLAARERALKESSEGGTPO 1299
 QY 1336 PPER-----HSPH--HLKEQHHR-----GSITQGI 1359
 DB 1300 PPRPPSPRYDAPPTLHHHSPHSHARHEPVRLWGDPPARRELGYRAGLSQEKAL 1359
 QY 1360 PRSYVEAQEDYLRR-----EAKLKREGTPPPPPSRDLTEAYKTOALGPKLKPAHE 1412
 DB 1360 TASPPAARSLHLRLPPTAPGVGPLLQGL-PEPTPHGVSKAWRTAA----- 1407
 QY 1413 GLVATVKEAGRSIHIEIPREELRHTE-IPLAPRPLKEGSITQGTPLKVDGTASTGSKKH 1471
 DB 1408 -----PEE-----PERLPHVRFLN-----CQARP-----PPAGTRGSSTE 1439
 QY 1472 DVRSLLIGSPGRTPPVHPLDMADARALERCYBESLSKSRPOTASSSGSARGAPVVP 1531
 DB 1440 D-----GPG--VPPPSPRVL-----FTSPSPGNEENGLPLIV- 1472
 QY 1532 ELGKPROSLTYEDHGAFFAGHLPRGSPVTWREPTRLQEGSLSSSKASQDRKLTSTPRE 1591
 DB 1473 ---LPPAPSPVDVDDG-----EFLFAEPLPPPLEFSNGFEK----- 1505
 QY 1592 TAKSPHSTVPEHHHPHPISEYHLLRGVGVLYRSHIPLAFDPTSIPIGILDAAYYL 1651
 DB 1506 ---PESPLTPGPHPLP-----DPPS--PATFLPAA----- 1531
 QY 1652 PRHLAPNTYPLHYLYIRGYPDTAALENRQTIINDYITQOMHNTATAMAQADMLR 1711
 DB 1532 -----PPPAVAAPPT--LDSTASSLSY----- 1553
 QY 1712 GLSPRESSLANYAAGPRGIIDLQOVHPLFVLVPTPTGTATAMDRLAYLTAPQPFSSR 1771
 DB 1554 -----DSEVATLTQGAAPAGD-----PPAFGPAPA---APAPPAPQF----- 1589
 QY 1772 HSSPLSPGGGTHLTKPTTTSSSERDRDRERDREREKSILTSITTTVEHAPWRPCT 1831
 DB 1590 -----GP---DPPPGTDSGIEEVDRSSSDHPLF-----TISSASTLSSL-----SA 1628
 QY 1832 EQSSGSSGSSGGGSSGRSPASHAHQHSPISPRTDALQORPSVLHNTGMKGIIITAVE 1891
 DB 1629 EGGNGTGVAGGGAGVAN-----GTLLDTYVAYLDGQAFGGSGTGP 1671
 QY 1892 PSKFTVL-----RSTSTSPVRPAATFPPTHCPCLGGLDGVYPTLMEPVLLPKAPR 1944

Db 1672 PYPQLMTPSKURGRALGTSGNLRPG-----PSGGLRDEVTPT-----SPT 1712
QY 1945 VARPERPRADTGHAFKAPARGLPSPKSGSEPRPLVPVSGHATARTPAKNLAP 2004
Db 1713 VS--VTGAGTDGLLALSACGPGTAGVAGP---VAVEVEVPVPLPA--ASLPRKLLP 1765
QY 2005 HHASPPPPAP--PASASDPHREKTQS-KPFSIQELRLSLGHSYSPGVEPVSPVS 2061
Db 1766 WEEGPGPPPPPLPGPLSQQASALATVKASIIISLSSKLQQFCGSS-TAGGALPWARGGS 1824
QY 2062 PSLT--HDKG---LPHLBEELDKSHLEBELRPQPGPVKLGGEAAHLPHLPLPESQPS 2116
Db 1825 GGSTDSHGCGASYIPERTSLQRLSED-----SQTSL-----SKPSS 1864
QY 2117 S-----PLQTAGPVKGHORVVTIAQHSIVITDYTEHHPQQLSAPLPAPLYSF 2166
Db 1865 SIFQWPKPLPLPLPTGSGVSSSTAAPGATSPASASASTR-HLQGVFEMRPL--- 1920
QY 2167 PGASCPVLDLRRPPSLYLPDPHGAPARGSPHSEGGKSPENKTSV----- 2214
Db 1921 -----LRRAPSPSLPASDH-----KVPAPRPSLPLPSGPIYPGL 1958
QY 2215 -----LGGEGDIEPV-SPPE-----CMTEPGHSRANVYPLLRYDGEQTEPSR 2256
Db 1959 FDIRSPTGAGGSTDPFAPVFPVPHFGISGGLGALSGASRS-----LSPTR 2006
QY 2257 MGSKSPGNT--SQPPAPFSK 2274
Db 2007 LLSLPPDKPFGAKPLGFWTK 2026

RESULT 76
ID ABO44406
AC ABO44406;
DT 26-SEP-2003 (first entry)
XX Novel human protein kinase #26.
XX Human; kinase; enzyme; cosmetic application; nutraceutical application.
XX Homo sapiens.
XX US6541252-B1.
XX 01-APR-2003.
XX 14-MAY-2001; 2001US-00854856.
XX 19-MAY-2000; 2000US-0206015P.
XX (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Hilbun E, Donoho G, Turner CA;
XX WPI; 2003-575927/54.
XX N-PSDB; ACH03789.
XX New nucleic acid encoding novel human proteins, useful in cosmetic and
XX nutraceutical applications.
XX Disclosure; Page; 11pp; English.
XX The invention relates to a new isolated nucleic acid encoding a novel
XX human protein kinase. The nucleic acid is useful in cosmetic and
XX nutraceutical applications. The present sequence represents the amino
XX acid sequence of a novel human protein kinase. Note: The sequence data
XX for this patent did not form part of the printed specification but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=6541252B1

XX SQ Sequence 2157 AA;
Query Match
Best Local Similarity 19.8%; Score 422; DB 7; Length 2157;
Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;
QY 147 PVSPPPHPTPELELVPPRLSKEELIQNMDRVREITVWEQIQISKKKQQOLEEAAK 206
Db 39 PLSLPQPSIPAAVPOSAPPPEPHEETV-----TATATSQVAQPPAAAPAGQAVA 89
QY 207 PREP-----EKPSPPPIESKRSILVQIYDENRKAEAHRLIEGLGQVDELPLY 257
Db 90 GPAPSTVPSSTSKDRPVSPSL-----VGSKEBPPPA 121
QY 258 NQPSDTRQYHENIKINQAMRKLLILYKRRNHARKWKOKFCORYDQLEALKEKKVERIE 317
Db 122 RSGS-----GSAKEPQERSQQDDI-EEETKAVGMS 155
QY 318 NNPRRAKESKV-REYEEKOFFEIRKOR-----ELQERMOSRVQQRGSLMSAARSE 369
Db 156 NDRFLKFDIEIGRSGFTVKGLDTETTVEVAWCELQDRK-----LTKSERQRF 205
QY 370 HEVSEIIDGLSQENLEKQMRQLAVIPMLYDADQORIK-----FINNGLM-ADPMKYK 424
Db 206 KEAEMLKGL-QHPNIVR-----FYDSWESTVKGKCIVLVTELMTSGTLKYL 253
QY 425 DR-----QVMNMWSEQEKETPREKFMHPK-----NFGLIA 455
Db 254 KRFKWKIKVLSWCQIILKGLQFLHTRTPPIIHRDLKCDNIFITGTSVKIGDLGL-- 311
QY 456 SFLEKRTVAECVL-----YYLTKNENY-----KSLVRRSYRRRKSQQOQQQQQ 501
Db 312 ATLKGASFAKSVIGTPEFWAPMEYBEKYDSDVDVAFGCMLEMATSEVPYSECQAAQI 371
QY 502 QOQQOQQOQP-----MFRSSQ-----EKDEKEKEKEKEKEKEKEKEKEKEDLLK 547
Db 372 YRRVTSQVGPASFDKVAIPEVKEIIEGCIQONKDERYSIKDLNHAFFQETGVRVELAE 431
QY 548 EKTDDTSGEDNDEKEAVASKRKTANSQGRKRITRSMANEANSE-EAITPOQSAELAS 606
Db 432 E-----DGEKIAIKLWIRIEDIKLKGKYKDNEAIFSFLEDRDVPDVAQ--- 478
QY 607 MELNESSRTEBEMETAKKGLLEHGRNWSAIAARMVGSKTQVQCKNFYFNFKKQNLDEIL 666
Db 479 -EMVESGYVCEGDHKTMAKIKDR-----VSLIK-----RKREORQLVR 516
QY 667 QOHLKMEKERNARKKKA-----PAASEEAAPPVVEDEMEASG 709
Db 517 EEQEKKEQESSLKQOQVEQSSASQTGIKOLPSASTGIPTASTTSASVSTQVEPEPEA-- 574
QY 710 VSGNEEVEEAALHASGNEVPRGCGSPATVN--NSSDTEISIPSPHTEAAKDTQNGP 767
Db 575 -DOHQOLYQQPSISVLSDGTVDGCGSVFTESRVSSQCTVSYGSH-EQAHTGTVP 632
QY 768 KPPATLGADGPPPG--PPT-----PPRRTSRAPTEPT-PASEATGAPTPPPAPPS 814
Db 633 HIPSTVQAQSPHGVPVPSVQGIQQTAPPQCTVQVLSQTSSEATTA-----QPV 687
QY 815 PSAPPPVPVKEEKEEETAAAPVE--EGEOKPAAAE--LAVDTKASEBPVSKSETEEA 870
Db 688 QFOAPQVLPQVAGKOLPVSPVPTIQGEPQIPVATQPSVVPVHGAFLPV----- 739
QY 871 BEGPAKGDAEAAEATAEGALKAEKGGGRATTAKSSGAPODSSATCSADEVDEAE 930
Db 740 -----CQPLTPPL-----LPQYVPSQIPISIPHVSTAQ 767
QY 931 GGDKNRLLSPRSLTPTGDPANASFPKPLDLKQLKQAAAAIP-----PIQVTKVHEP 984
Db 768 TG-----FSSLPTWA-----AGITQPLLTASSATTAIPGVSTVVPVSLTLQ 814
QY 985 PREDAAPTKAPAPPAPPNOLQESDAPQOQSSPRGKRSRPPADKAEFAAEAKLPG 1044

815 VTQ-----LFSQVHPQLQP-----AVQSMGI-----PANL-GQAAEVLSSG 851
1045 DPPCWTSGLPVPPVPPREVIVKASHPADPSAFSYPBPGHPLPLGLDHTARPVL--DRPPTI 1102
852 D---VLYQGPPLRPPOY-----PGSNIAPSSNVASCIHST---VLXPPMPTEV 896
1103 SNPPPLISSAKHPSV---LEROIGALISQMSQVQLHVPYSEHAKAPVGPVTMGILPLPMDPK 1159
897 LATPGYFTVQVYVESNLVPMGGV--GGQVQSQPGSLAQAPITSSQOAV-----947
1160 KLAPSGVQEOQLSPRQAGPDESIGV-----PTAQEASVLRGALGVPGSGITKGI 1212
948 -LESTQGV-----SOVAPAEVAVAOQATPTTLASSV--DSAHSDVASG-MSDG- 994
1213 PSTRVPSDAITYRGSI THGTADVLKYGTIRIIGEDSPSLDRGREDSLPKGHVIEG 1272
995 -NENVPSSG-----RHEGRTRKHRYKSVRSRSHKTSRPLKRLINVS 1038
1273 KKGHVLSYEGGMSVTOCSKE-----DGRSSGPPHETAAPKRTYDMMEGRVG 1319
1039 NKGD-----RVVEQCLETHNRKMVTFKDL DGN-----PEEIAIWNNDFI-----1081
1320 RAISSASIEGLMGRALPPRHSPHLLKQOHHIRGSI TQIGIPRSYVQAQEDYLRRRAKLJK 1379
1082 LAIERESFVDQVREII---EKADEMLSEDSVSEPEGDQGL--ESLQKDDYGFSGSKLE 1136
1380 RSGTTPPPPPSRDLTEAYKTQALGPLKLKPAHEGLVATVKEAGRS--IHEI PREELRHTP 1437
1137 GFQKQIPASSM-----POQIGIPTSSLQVHVHSGARRFVSPVESRLRESK 1184
1438 ELFLAPRLKEGSI TQGTPLKYDTGASTTGSK-----KHDVRSLL-----1476
1185 VFP-----DTVAASTAQSPGMNLSHASSLSLQQAFLSELRAQMT 1229
1477 -----IGSPQRTPPVHPDLMADARALERACYEBSLKSRPGTASSGSGSIARGAP- 1527
1230 GNTAPPNFSGTGPFPVVP-----FLSSIAGVPTTAAATAPVATSP 1275
1528 -----VIVPELCKPRQSL--TYEDHGAFFAGHLPRGSPVTMREPTPLRQSGSLSSKA 1579
1276 NDISTSVIOSEVTVTEEGIAGVATSGVTGSGL-----PIPPVSESPVLSVV 1325
1580 SDRKLTSTPREIA---KSPHSTVPEHHPHPISPYEHLLRGVGDLYRSHIPLAPDPTS 1636
1326 SS-----ITIPAVVSISTSPSLQVPTSTSEIV-----VSSITALYPS-----1362
1637 IPRGIPLDAAYLPHRIAPNPTYPHLYPPYLI-----RGYPT 1676
1363 ----VTVSATSASAGGSTATPGK-----PPAVSQQAAGSTTVGATLTSVSTTSPST 1413
1677 AALENRQ-----TIINDYITSQMHNTATAMA-----QRA 1707
1414 ASQLSLSSTSTPTLSTVVSASHLSLTKSHSSTGLAFSLASAPSSSSPCAGVSSVI 1473
1708 DMLRGLSPR-----ESSIALNYAAGRGIIDLSQVPHLPVLVPPPTGPTATAMRLAYL 1761
1474 SQPGLHPLVPSIVIASTPILQAAGSTPLLPQVPSIPLVQPVANPAV--QQTLL 1530
1762 PTAPQPFSSRHSSPLSPGCP--THLTKPTTSSSERDRDRDRDREREKSLTSTTT 1820
1531 HSQQPQ-----ALLPNQPHTHCP-----EVDSDTQPKAPGIDDIKT 1566
1821 VEHAPIWRPCTQSGSSGSGSGSGSSRSPASHSHAHQHSPISTPTQDALQRPVSLHN 1880
1567 LEE-----KURLSEHSSGA-----QHASVSLTS-----LVIES 1598
1881 TGMKGI-ITAVBPSKPTVLRLSTSSVPRPAATFPFATHCPLGGTLDGVYPTIMEFVLLP 1939
1599 TVTPGIPTTAVAPSK--LLTSTT-----STCLPPTNLPLG-----1631
1940 KEAPRVARERPADTGHAFLAKPPARSGLEPASPSSKSEPRPLVPPVSGHATARTPA 1999
1632 ----TVALPVTVPVTPGQVSTPVSTTTSVGKPGTAPSKPPLTKAPVLPVGTLPAGTLES 1687

QY 2000 KNLAPHASPDPPAPPASADSPHREKTS--KPFSTQELRLSLGYHGSYSPEGVPEVSP 2058
Db 1688 EQL-----PPFGPSL-----TQSQPLEDLDAQLR-----RTLSEXTIVTSA 1726
QY 2059 VSSPSLTHDKGLPKHLEELDKSHLEGLRPKQGPVKLGGEAAHPLRLPLPE---SQPS 2115
Db 1727 V-----GPVMAAPTA-ITEAGTQPKQGVSVK 1753
QY 2116 SSPILQTAGV---KGHQRVVTLAOHISEVITQDTRHHPOOLSAPLPAPLYSPGASC 2171
Db 1754 EGFVLATSSGAGVFKVGRFOVSVA--DGAQKEGKNKSEDAKSVHFESTSES 1804
QY 2172 PYLDLRRPPSLDLYLPPP-----DHGAPARGSPHSEGGKRGSPENKTS-----2213
Db 1805 SVLSSSSPESTILVKEPENGITIPGISSDVPESAHKTTASEAKSDTGQTKVGRFOVTTTA 1864
QY 2214 -----VLGGGEGDIE-----PVSPPEGMTPEGHSRAVYPLLRYDGEQTPPSRMGSKS 2261
Db 1865 NKVGRFSVSKETDKITDKKEGFPVASPPFMDLEQAVLPAVLPKKEKP-ELSEPSHLN---1920
QY 2262 PGNTSQP-PAPFSKLTESANMVKSKQEIKNKLNHNREPEYNIQSPTGIFNMPAIT 2320
Db 1921 -GPSSDPEAAFLSRDVGSGSPHSPHQLSSKSL-----PSQNLSSLSNSFNSSYMS 1972
QY 2321 GTGLMTYRSQAVO-----EHAGTNMGLEAIIRKALMGKYDOWEESPPLSANAFNPLN 2372
Db 1973 SNESDIEDLKLRLRLRDKHLKEIQDLSQKHEIESLYTKLGKVP-----2022
QY 2373 ASASLPAAMPITAAOGRSDHTLTPSGGGKAKVSGRPSRKAQSP-APGLASGDRPPSV- 2430
Db 2023 -AVIIPPAAPLS---GRRRRPTKS-----KGSKSRSSSLGNKSPQLSGLNLSQQAASVL 2073
QY 2431 ---SSVHSEGD 2438
Db 2074 HPQOTLHPGN 2084
RESULT 77
ABO44390
ID ABO44390 standard; protein; 2217 AA.
XX ABO44390;
XX AC ABO44390;
XX DT 26-SEP-2003 (first entry)
XX Novel human protein kinase #10.
XX Human; kinase; enzyme; cosmetic application; nutraceutical application.
XX Homo sapiens.
XX OS Homo sapiens.
XX PN US6541252-B1.
XX PD 01-APR-2003.
XX PF 14-MAY-2001; 2001US-00854856.
XX PR 19-MAY-2000; 2000US-020601SP.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Walke DW, Hilbun E, Donoho G, Turner CA;
XX WPI; 2003-575927/54.
XX N-PSDB; ACH03773.
XX PT New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.
XX PS Disclosure; Page; lipp; English.
XX CC The invention relates to a new isolated nucleic acid encoding a novel

CC human protein kinase. The nucleic acid is useful in cosmetic and
CC nutraceutical applications. The present sequence represents the amino
CC acid sequence of a novel human protein kinase. Note: The sequence data
CC for this patent did not from part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6541252B1
XX
SQ Sequence 2217 AA;

Query Match 3.2%; Score 422; DB 7; Length 2217;
Best Local Similarity 19.8%; Pred. No. 1.8e-14;
Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;

QY 147 FVSPSPPHPTDELELVPRLKEELIQNMDRVDREITVMEQISLKKKKQQLLEBAK 206
DB 99 PLSLPQSPAPAAVQSPAPPEHRETV-----TATATSQVAAQPPAAAPGEQAVA 149

QY 207 PPEP-----EKPVSPPPIESKHSVLQIYIDNRKKAABAHRIELGLGPVELPLY 257
DB 150 GPAPSTVPSSTSKDRPVSPSL-----VGSKEPPPPA 181

QY 258 NQPSDTRQYHENIKINQAMRKLILYFERNHARKQWKFCQRYDOLMEALEKKVERIE 317
DB 182 RSGSG-----GGSKEPQERSQQDDI-EELLETKAVGMS 215

QY 318 NNPRRRAKESKV-REYIEKQFPPIRQR-----ELQERMOSRVQQRGSLGMSAARGE 369
DB 216 NDRFLPKFDIEICRGSKFYVYKGLDTTVEVAVCELDQRK-----LTKSERQRP 265

QY 370 HEVSEIIDGLSGOENLEKQMRQAVIPMLYDADQRIK-----FINNGML-ADPMKVYK 424
DB 266 KSEAEMKGL-OhpNIVR-----FYDSWESTVAKKCVILVTELMTSGTLKTYL 313

QY 425 DR-----QVMNWSQEKETPREKFWHPK-----NFGLIA 455
DB 314 KRPFKWKIKVLRSWCQIILKGLQFLHTRTPPIIHRDLKCDNIFITGTSVKIGDLGL-- 371

QY 456 SFLEKRTVAECVL-----YYLTKNENY-----KSLVRRSYRRRGKSOQOQOQOQ 501
DB 372 ATLKRASPAKSVIGTFEFWAPMEYKEDVDVYAFGCMLEMAISEYVSECQNAQI 431

QY 502 QOQOQOQOQO-----MPRSSQ-----EKKDEKEKEKAEKEKEKEPEVENDKEDLLK 547
DB 432 YRRVTSGVKSPASDKVAIPEVKEIIEGCIRQNDKERYSIKDLNHAFFQOETGVRVELAE 491

QY 548 EKTDDTSGEDNDEKAEVASKRTANSQGRKRGRITSMANEANSE-EALTPOQSAELAS 606
DB 492 E-----DDGEKIAIKLWRIEDIKKLGKDYKDNKAEIEFSDLERDVPEDVAQ--- 538

QY 607 MELNERRTEEMETAKKGLLEHGRNWSAIRMVSGSKTVSQCKNFYFNYKKRQNLDEIL 666
DB 539 -EMVESGYCEGHDKTMAKAIKDR-----VSLIK-----RKREORQLVR 576

QY 667 QQHLKMEKERNARRKKKA-----PAAASEEAPPPVPEDEMEASG 709
DB 577 EEQKKKQESSLKQVQESSASQGIKQLPASASTGIPTASTTSASVQTVPEEPEEA-- 634

QY 710 VSGNEEMVEEAEALHASGNEVPRGCSGPATVN--NSDTEIPSPHTEAAKDTQNGP 767
DB 635 -DOHQLOLQOQPSISVLSGTVDSGGSSVFTESRVSSQOQTVSYGSQH-EQAHSTGTVP 692

QY 768 KPPATIGADGPPPG--PPT-----PPRTRAPTEPT-PAGEATGAPTPPPAPPS 814
DB 693 HPESTVQASQPHGVVPPSPVQOQIQOATPQQTQVYSISQTSSEATTA-----QPVS 747

QY 815 PSAPPVVPVKEKEEBETAAPPVE--EGEQKPPAAEE--LAYDTGKAEPPVKECTEEA 870
DB 748 QPQAPQVLPQVSAKQLPVSQPTTIGEPQIPVATQPSVVPVHSGAHFLPV----- 799

QY 871 EEPGAKGDAEAAEATAGALKAEKGGSGRATTAKSSCAPQDSSSATCSADEYDEAE 930
DB 800 -----GQPLTPL-----LPQYVPSQIPISTPHVSTAQ 827

QY 931 GGDKNRLLSPPRSLLTPTGDDPRANASPOKPLDLKQLKQRAAIP-----PIQTVKHBP 984
DB 828 TG-----FSSLPTTMA-----AGITQPLLTASSATTAIPGVSTVVPSPQLTLP 874

QY 985 PREDAATKPAAPPPPPONLOPESDAPQOQSSPRGKSRSPAPPADKAEFAAEAKLPG 1044
DB 875 VTQ-----LPSQVHPQLQP-----AVQSMGI-----PAVL-GQAAEVLSSG 911

QY 1045 DPFCWTSGLFPFVPPREVIKASPHAPDPDSAFYAPPGHPLPLGLHDTARPVL--PRPPTI 1102
DB 912 D--VLYQGFPPPLPQY-----PGDSNIAFSSNVASVCIHST--VLXPPMPTEV 956

QY 1103 SNPPPLISAKHPV--LEROIGALSQMSVOLHVYPYSEHAKAPVGPVTMGLPLPMDPK 1159
DB 957 LATPGYFTVQVQVYESNLLVPMGV--GGQVQVSPGGSLAQAPTSSQAV----- 1007

QY 1160 KLAPSGVKQEQLSRPGQAGPPESLGV-----PTAQEASVLRGTALSGVPGGSLTKGI 1212
DB 1008 -LESTQGV-----SQVAPAEVAVAQOQTPTTLASSV--DSAHSDVAG-MSDG- 1054

QY 1213 PSTRVPSDSAITYRGSITHGTADVLTKGTRIIGEDSPSRDLRGREDSLPKGHVIEG 1272
DB 1055 -NENVPSSG-----RHEGRTTKHYRKSVRSRSEKTSRPKLRILNVS 1098

QY 1273 KXGHVLSYEGGMSVTQCSKB-----DGRSSGPPHETAAPKRTYDMMGRVG 1319
DB 1099 NKGD-----RVVECOLETHNRKMVTFKFLDGDN--PBEIATIMVNDFI----- 1141

QY 1320 RAISSASIEGLMGRAIPEPHSPHLKEQHHRIGSITQIGIPRSYVEAQEDYLREAKLLK 1379
DB 1142 LAIERESFVDQVREII--EKADEMLSESDVSEPEGDQGL--ESLQKDDYFGSGQKLE 1196

QY 1380 REGTTPPPPPRDLTEAYKTQALGPLKPAHEGLVATVKEAGRS--IHEIPREERLHTP 1437
DB 1197 GEFKQPIPASSM-----PQOIGIPSSLTQVHVSAGRRRIVPVPESRLRESK 1244

QY 1438 ELPAPRLKESITQGTPLKYDTGASTGSK-----KHDVRSI----- 1476
DB 1245 VFP-----DTVAATAQSPGMNLSHASSLSLQOAFSELRAQMT 1289

QY 1477 -----IGSCRTFPPVHPLDMADARALERACYEESILKSRPGTASSSGSIARGAP- 1527
DB 1290 GNTAPNFSHTGPTFPVVPP-----FLSIAQVPTTAATAPVATSSPP 1335

QY 1528 -----VIVPELCKPRQSL--TYEDHGAPFAGHLPGRGSPVTMRPTRLQEGSLSSKA 1579
DB 1336 NDISTSVIOSEVTVTEEGIAVATSTGVVTSGL-----PIPPVSESVLSSVV 1385

QY 1580 SQDRKLTSTPREIA--KSPHSTVPBHHPHIPISPYEHLRGVSGVDLYRSHIPLAFDPTS 1636
DB 1386 SS-----ITIPAVVSVISSTSPSLQVPTSTSEIV-----VSSITALYPS- 1422

QY 1637 IPRGIPLDAAYAYLPHLAPNPTYPHLYPPYLI-----RGVYDPT 1676
DB 1423 ---VTVSATSASAGGSTATPGK-----PPAVVSQQAAGSTTVGATLTSVSTTSPST 1473

QY 1677 ALENRQ-----TIINDYITSQOMHNTATAMA-----QRA 1707
DB 1474 ASQSLQLSSSTSTPTLAETVVVSAHSLDKTSHSSTGLAFSLASAPSSSSPGAGVSSYI 1533

QY 1708 DMLRGLSPR-----ESSIALNYAAGRGIIDLSQVPHLPVLVPPTEGTATMDRLAYL 1761
DB 1534 SQPGLHLPLVIPSIVIASTPILQOAGPTSTPLLPQVPSIPPLVQPVANVAV--QOTLI 1590

QY 1762 PTAPQPFSSRHSSPLSPGCP--THLTKPTTSSSERERDRDRDREREKESILSTTT 1820
DB 1591 HSQOPQ-----ALLPNQPHTCP-----EVSSTQPKAGIDDIKT 1626

QY 1821 VEHAPTRPQTESGSGSGSGSGSGSGSRPASHASHAHQHSPISTQDALQORPSVLHN 1880
DB 1627 LEE-----KLRLSEHSSSCA-----QHASVLETS-----LVIES 1658

QY 1881 TGMKGI-ITAVEPSKPTVLRSTSTSTSFVRPAATFPFATHCPLGGLDGVVPTLMEPVLLP 1939

Db 1659 TVTPGIPTTAVAPSK--LLTSTT-----SCLPFTNLPLG----- 1691

Qy 1940 KEAPRVARPERPRADTGHAFLAKPPARSGLEPASSPSKSGEPRLVPPVSGHATARTPA 1999

Db 1692 ----TVALPVTVVTVQGVSTPVSTTSGVKGTAFSKPPLTKAPVLPVGTLPAGTLP 1747

Qy 2000 KNLAPHASPDPPAPPASADPHREKTQS--KPFISOELRLSLGYHSGSYSPGVPVSP 2058

Db 1748 EQL-----PPFPGPSL-----TQSOQPLEDLDAQLR-----RTLSPEXIVTSA 1786

Qy 2059 VSSPSLTHDKGLPKHLELDKSHLEGELRPQGPVKLCGEAAHLPHLPLPE---SQPS 2115

Db 1787 V-----GPVSMAPTA-ITEAGTQPKGVSQVK 1813

Qy 2116 SPSLLQTAGV---KGHORVVTLAQHISEVITQDYTRHHPPQOLSAPLAPLYSFGASC 2171

Db 1814 EGVLATSSGAGVKGVRGFQSVAA-----DGAQKGNKSKEDAKSVHFESTSES 1864

Qy 2172 PVLDLRRPSSDLYLPPP-----DHGAPARGSPHSGGKRSPEPNKTS----- 2213

Db 1865 SVLSSSSPSTLVKPEPNGITIPGISSDVPESAHKTTASEAKSDTGQPTKVGRFQVTTTA 1924

Qy 2214 -----VLGGEDGIE-----PVSPEGMTPEGHRSNAVYPLLYRDGQTEPSRMGSKS 2261

Db 1925 NKVGRFSVSKTEDKITDTTKKGVASPPFMDLEQAVLPAPVPKKEKP-ELSEPSHLN--- 1980

Qy 2262 PGNTSOP-PAFTSKLTFESAMVSKKOEINKLNTNREPEYNIQSOPGTIENMPAIT 2320

Db 1981 -GPSSDPEAFUSRDVDDGSGSPHQSLSKSL-----PSQNLQSLSNFSNFSYMS 2032

Qy 2321 GTGLMYRSQAVQ-----EHAFTNMGLEAIIRKALMGKYDQWESPPLSANAFNPLN 2372

Db 2033 SDNESDIEDLKLRLRLDRKHLKEIQDLQSRQKHEIESLYTKLGKVP----- 2082

Qy 2373 AGASLPAAMPITAADORSHTLTSPPGGGKAKVSGRPSRKAQSP-APGLASGRDPPSPV- 2430

Db 2083 -AVIIPPAAPLS---QRRRRPTKS-----KGSKSSRSSSLGNKSPQLSGLNLSQGAASVL 2133

Qy 2431 ---SSVHSEGD 2438

Db 2134 HFQQTLLHPGN 2144

RESULT 78

ID ABO44405

AC ABO44405;

DT 26-SEP-2003 (first entry)

XX Novel human protein kinase #25.

DE Human; kinase; enzyme; cosmetic application; nutraceutical application.

KW Homo sapiens.

OS US6541252-B1.

PN 01-APR-2003.

PD 14-MAY-2001; 2001US-00854856.

XX 19-MAY-2000; 2000US-0206015P.

PR (LEXI-) LEXICON GENETICS INC.

XX Walke DW, Hilbun E, Donoho G, Turner CA;

PI WPI: 2003-575927/54.

DR N-PSDB; ACH03788.

XX

PT New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.

XX Disclosure; Page; 11pp; English.

XX The invention relates to a new isolated nucleic acid encoding a novel human protein kinase. The nucleic acid is useful in cosmetic and nutraceutical applications. The present sequence represents the amino acid sequence of a novel human protein kinase. Note: The sequence data for this patent did not from part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6541252B1

XX Sequence 2294 AA;

SQ

Query Match 3.2%; Score 422; DB 7; Length 2294;

Best Local Similarity 19.8%; Pred. No. 1.9e-14;

Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;

Qy 147 PVSPTSPHTDPELELVPRLSKEELIQNMDRVREITVVEQQISKLKKQQQLEEEAAK 206

Db 39 FLSLPQPSIPAAPVQSAPEPEHRETV-----TATATQVAQPPAAAPGEQAVA 89

Qy 207 PPEP-----EKVPSPPIESKHSRLVQIYDENRKKABAAHRIEGLQGPVELPLY 257

Db 90 GPAPSTVPSSTSKDRPVQPSL-----VGSKEEPPPA 121

Qy 258 NQPSDTRQVHENIKINQAMKKLILYFKERNHARKQKFCQRYDQLEALEKKVERLE 317

Db 122 RSGSG-----GSAKEPOEERSQQDDI--EELTKAVGMS 155

Qy 318 NNPRRRAKESKV-REYIEKQFPEIRKQR-----ELQERMQSRVQGRSGISMSAARSE 369

Db 156 NDRFLKFDIEIGRSPKTVYKGLDTETVEVAVCELQDRK-----LTKSERQRP 205

Qy 370 HEVSIHIDGLSQENLEKQMLAVIPWLYDADQORIK-----FINMGLM-ADPMKVYK 424

Db 206 KEAEAMLKGL-QHPNIVR-----FYDSWESTVTKKKCIVLVTELMSTGTLKTVL 253

Qy 425 DR-----QVNMWSEQSKETFRKFMQHPK-----NFGLIA 455

Db 254 KRPVKMKIKVLSWCRQILKGLQFLHTPTPIIHRDLKCDNIFITGPTGSKVIGDLGL-- 311

Qy 456 SFLEKRTVAECVL-----YYILTCKKNY-----KSLVRRSYRRRGSKSQQQQQQQ 501

Db 312 ATLKRASFAKSVIGTPEFMAPEMYEKEYDESVDVYAFGCMLEMATSEVPYSECNAAQI 371

Qy 502 QQQQQQQQQP-----MPRSSQ-----BEKDEKEKEKEAEBEKEPEVENDKEDLLK 547

Db 372 YRRVTSGVKPASFDKVAIPEVKEIIEGCTRONKDERYSIKOLLNHAFFQEEETGVRVELAE 431

Qy 548 EKTDDTSGEDNDEKAVASKGRKTANSQGRKGRITRSMANEANSE-EAITPQQSAELAS 606

Db 432 E-----DDGEKIAIKLWRIEDIKKLQKGYKDNEAIEFSDLRDVEDVAQ--- 478

Qy 607 MELNESSRWTEEMETAKKGLLEHGRNWSAIRMVGSKTVSQCKNFYFNKQRQLMDEIL 666

Db 479 -EMVESGYVCEGDHKTMAIKDR-----VSLIK-----RKREQQLVR 516

Qy 667 QQHKLKMEKERNARKKKA-----PAAASEEAAFPVVEDEMEASG 709

Db 517 EQEKKQKESSLKQOQVEQSASQTGIKQLPSASTGIPASTTSASVSTQVPEPEA-- 574

Qy 710 VSGNEEEMVEEAALHAGNEVPRCEGSGPATVN--NSSDTESIPSPHTEAKDTGONGP 767

Db 575 -DQHQQQLYQQPSISVLSDGVDSQSSVFTESVSSQQTVVSQSH-EQAHSTGTVP 632

Qy 768 KPPATLGADGPPPG--PPT-----PPRRTSRAPIEPT-PASEATGAPTPPPAPPS 814

Db 633 HIFSTVQAQSPHGVYFPSSVQOQIQQTAPPOQTQVYSLSQTSSTSEATTA-----QPVS 687

Qy 815 PSAPPVVPKKEEETAAAPVE--EGEEQKPPAAER--LAVDTGKAEFPKSECTEEA 870

Db 688 QQAQOVQLQVSGAGKQLPVSPVPTTQGEPOIPVATOPSPVVPVHGAHFLPV----- 739
QY 871 BEGPAKGDAEAAEATAEGALKAEKKEGSGRATTAKSSCAPQDSDSSATSCADEYDEAE 930
Db 740 -----GQPLPTPL-----LQPVQVPIPISTPHVSTAQ 767
QY 931 GGDKNLLSPRSLTPTGDPANASPOKPLDLKQLKQRAAAIP-----PQVTKVHEP 984
Db 768 TG-----FSSLPTMA-----AGITQPLTLASSATTAALPGVSTVVPVQLPTLQP 814
QY 985 PREDAAKTPAPPAPPPQNLQSPESAPQPGSSPRGKSRSPAPPADKEAFAAEAQKLPQ 1044
Db 815 VTQ-----LPSQVHPQLQP-----AVQSMGI-----PANL-GQAAEVLPSG 851
QY 1045 DPPCWTSGLPFPVPPREVIKASPHADPPSAFSAFPGHPLPLGLHDTARPVL---PRPPTI 1102
Db 852 D---VLXQGFPPRLPPQY-----PGDSNIAPSSNVAVCIHST---VLXPPMPTEV 896
QY 1103 SNPPPLISSAKHPSV---LERQIGALSQMSVOLHVPYSEHAKAPVGPVTMGLPLPMDPK 1159
Db 897 LATPGYFPTVQPVYBSNLLVPMGGV---GGQVQVSPFGGSLAQAPTSSQQA----- 947
QY 1160 KLAPSGVQEQQLSPRGQAGPPESLGV-----PTAQASVLRGTLGSLVPGGSITKGI 1212
Db 948 -LESTQGV-----SQVAPAEVVAQPOATQPTTLASSV---DSAHSDVASG-MSDG- 994
QY 1213 PSTRVPSDAITVRSITHTGTPADVLVYKGTITRIIGEDSPSLRDGRDLSLPKHVYIEG 1272
Db 995 -NENVESSG-----RHEGTTKXHYKSVRSRSRHEKTSRPKLRILNVS 1038
QY 1273 KXGHVLSYEGGMSVTCQSK-----DGRSSGPPHETAAAPKRTYDMEGRVG 1319
Db 1039 NKGD-----RVBECQLETHNRKMTVKFDLDGDN---PPEIATIMVNDFI----- 1081
QY 1320 RAISSASIEGLMCRAIPPERHSPHLKEQHIGSITQGIIPRSYVEAQEDYLREAKLK 1379
Db 1082 LAIERESFVDQVREII---EKADEMLSEDSVEPEQDQGL---ESLQKDDYGFSGSKLE 1136
QY 1380 REGTPPPPPSRDLTEAYKTAQALGPLKLPAHGLVATYKEAGRS---IHEIPREELRHTP 1437
Db 1137 GEFKQPIPASSM-----PQIGTPTSSLTQVHSAGRRTIVSPVPSRLRESK 1184
QY 1438 EULAPRPLKEGSIITGTLKYDTGASTTGSK-----KHVRSI----- 1476
Db 1185 VFP-----DTVAASTAQSPGMNLSHASSLSLQQAFAFSELRRQAMTE 1229
QY 1477 -----ICSPGRTFPVPHPLDMADARALACVYESLKSRCPTAGSSGSIARGAP- 1527
Db 1230 GPNAPPNFSHTGTPFPVPP-----FLSSIAGVPTTAAATAPVPATSSPP 1275
QY 1528 -----VIVPELKGPROSL---TYEDHGAPFAGHLPGRGSPVTMREPTPRLOEGLSSSKA 1579
Db 1276 NDISTSVIQSEVTVPEEGTAGVATGTGVTSGL-----PIPPVSESPVLSSVV 1325
QY 1580 SDRKLTSTPREIA---KSPHSTVPEHHPHPIPSYEHLLRGVSGVDLYRSHIPLAFDPTS 1636
Db 1326 SS---ITIPAVVSIITSPSLQVPTSTSEIV-----VSSTALYPS----- 1362
QY 1637 IPRGIPLDAANAYYLRHLAPNTYTHLYPPYLI-----RGYPDT 1676
Db 1363 -----VTVSATSSAGSGSTATPGPK-----PPAVVSOQAAGSTTVGATLTSVSTTSPFST 1413
QY 1677 AALNRQ-----TIINDYITSQMHNTATAMA-----QRA 1707
Db 1414 ASQLSTQLSSSTPTPLAETVVVSAHSLDKTSHSTTGLAFSLSAPSSSSPGAGVSSYI 1473
QY 1708 DMLRGLSPR-----ESSALNVAAGRGHIDLSQVPHLPVLVPPPTGTPATMDRLAYL 1761
Db 1474 SQPGLHPLVIPSVIASTPILPQAAGTSTPLLPQVPSIPLVQPVANVPAV---QQTLLI 1530
QY 1762 PTAPOFSSHSSSPSLSPGCP---THLTKPTTSSSERDRDRDREREKILSTTTT 1820
Db 1531 HSQFQP-----ALLPNQPTHCP-----EVDSDTOPKAPGIDDIKT 1566

QY 1821 VEHAPIWRPGTQSSGSSGSSGSSGSSSRPASHSHAHQHSFISPRTDALQORPSVLHN 1880
Db 1567 LEE-----KLRSLSFSEHSSGA-----QHASVSLETS-----LVIES 1598
QY 1881 TGMKGI-ITAVBPSKPTVLRSTSTSPVRPAATFPDPATHCPGLGTLGDGVVPTLMEPVLLP 1939
Db 1599 TVTPGILPTTAVAPSK---LUTSIT-----STCLPTNLPLG----- 1631
QY 1940 KEAPRVARPERPRADTGHAFKAPPARSGLEPASSPSKSEPRPLVPVSGHATIAARTPA 1999
Db 1632 ----TVALPVTVPVTPGVQVSTPVSITTSVGKGTAPSKPPLTKAPVLPGTELPACTLPS 1687
QY 2000 KNLAPHASDPDPAPPASASDEHREKTS- KFSIQEELRSLGSLGSSYSPGVPFVSP 2058
Db 1688 EQL-----PPFGPSL-----TOSQPLEDLIDAQLR-----RTLSPXIIWTS 1726
QY 2059 VSSPSLTHDKGLPKHLEELDKSHLEGELRPQGPVKLGGEAAHPLHPLPS---SQPS 2115
Db 1727 V-----GPVSMAPTA-ITEAGTOPKGVSOVK 1753
QY 2116 SSPLLQTAFGV---KGHQRVVTLAGHISEVITQDYTRHHPQQLSAPLPAPLYSFFCASC 2171
Db 1754 EGPVLATSSGAGVFKMGREFQVSA-----DGAQKEGKNKSEDAKSVHESSTSES 1804
QY 2172 PVLDRRRPSDLYLPPP-----DHGAPARGSPHSEGGKRSPEPNKTS----- 2213
Db 1805 SVLSSSSPESTLVKPPNGITIPGISSDVPESAHTTASEAKSDTGQTKVGRFQVTTTA 1864
QY 2214 -----VLGGGEGDIE-----PVSPEGMTPEGHSRSVYLLYRDGQTEPSRMGSKS 2261
Db 1865 NKVGRFSVSKTEDKITDTKKEGFPVSPFMDLEQAVLPVPIPKKEP-ELUSEPSHLN--- 1920
QY 2262 PGNTSOP-PAFTSKLTESNAMYKSKQIKNKLNTHRNEPEYINISQPGTEIFNMPAIT 2320
Db 1921 -GPSSDPEAAFLSRDVEDGSGSPHSPHQLSSKSL-----PSQNLSQLSNSFNSSYMS 1972
QY 2321 GTGLMYRSQAVQ-----BHASTNMGLEAIIRKALMGKYDQWEESSPLSANAFNPLN 2372
Db 1973 SONESDIEDLKLRLRLDKHLKEIQDLQSRQKHEIESLYTKLGKVP----- 2022
QY 2373 ASASLPAAMPITAADGSDHLLTSPCGGGKAKVSGRPSRKAASP-APGLASGDRPPSV- 2430
Db 2023 -AVIIPPAAPLS---GRRRRPTKS-----KGSKSSRSSSLGNKSPQLSGNLQQAASVL 2073
QY 2431 ---SSVHSEGD 2438
Db 2074 HPOQTLHPGN 2084

RESULT 79

ABO44389
ID ABO44389 standard; protein; 2354 AA.

XX ABO44389;

XX XX
DT 26-SEP-2003 (first entry)

XX Novel human protein kinase #9.

XX Human; kinase; enzyme; cosmetic application; nutraceutical application.

XX Homo sapiens.

XX US6541252-B1.

XX 01-APR-2003.

XX 14-MAY-2001; 2001US-00854856.

XX 19-MAY-2000; 2000US-0206015P.

XX (LEXI-) LEXICON GENETICS INC.

XX	Walke DW, Hilbun E, Donoho G, Turner CA;	1045	DPCHWTSGLPFPVPPREVIKASPHADPSAFSYPAGHPHPLPLGLHDTARPVL--PRPPTI	1102
PI	WPI: 2003-575927/54.	912	D--VLYQGFPPRLPOY-----PGDSNAPSSNVASVCIHST---VLXPPMPTVE	956
DR	N-PSDB; ACH03772.	1103	SNPPPLIISAKHPV---LEROIGAIQCMVSQVQLHPVSEHAKAPVGPVTMGLPLPMDPK	1159
DR	New nucleic acid encoding novel human proteins, useful in cosmetic and	957	LATPGYFTVTVQVYVESNLLVPMGV--GGQVQVSGPGSLAQAPTSSQAV-----	1007
PT	nutriceutical applications.	1160	KLAPSGVQKQOLSPRGAGPPESLGV-----PTAQEASVLRGTALGSPVGGSIITKI	1212
XX	Disclosure; Page: 11pp; English.	1008	LESTQGV-----SQVAPAEVAVAPQOATQPTTLASSV--DSAHSDVASG-MSDG-	1054
XX	The invention relates to a new isolated nucleic acid encoding a novel	1213	PSTRVPSDAITYRGSITHGTPADVLYKGTITRIIGEDSPSLDRGSDSLPKGHVIEG	1272
CC	human protein kinase. The nucleic acid is useful in cosmetic and	1055	NENVPSSG-----RHEGRTTKRHYKRSVRSRSHKETSRLKRLINVS	1098
CC	nutriceutical applications. The present sequence represents the amino	1273	KKGHVLSEYEGGMSVTCQSK-----DCRSSGPPHETAAPKRTYDMMEGRVG	1319
CC	acid sequence of a novel human protein kinase. Note: The sequence data	1099	NKGD-----RVVECCOLETHNRKMWTFKPDLDGDN---PEEITATIMVNDFI----	1141
CC	for this patent did not from part of the printed specification but was	1320	RAISSASIEGLMGRAPPERHSPHHLKEQHHRGSIITQIGIPRSYVEAEDYLRRRAKLK	1379
CC	obtained in electronic format directly from the USPTO at	1142	LAIERESFVDQVREII---EKADEMLSEDVSVEPEGDQGL--ESLQKDDYDFSGSKLE	1196
CC	seqdata.uspto.gov/sequence.html?DocID=6541252B1	1380	REGTTPPPPPSRDLTEAVYKQALGPLKPKAHEGLVATVKEAGRS--THEIPRELRHTP	1437
XX	Sequence 2354 AA;	1197	GEFKQIPASSM-----PQIGITPSSLTQVHSGARRFIVSPVPESRLRESK	1244
XX	Query Match	1438	ELPLAPRLKEGSIITQGTPLKYDTGASTTGSK-----KHDVRSL-----	1476
XX	Best Local Similarity	1245	VFP-----SEIT-----DTVAASTAQSPGMNLSHSASSLSLQQAFLSRRRAQWTE	1289
XX	Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;	1477	-----IGSPGRTFPVPHLDVMDARALERACEYESLKSRRPCTASSSGGSIARGAP-	1527
QY	147 PVSPPPPHPTDPELELVPPRLSKEELIONMDRVDRITWVEQOISKLKKKQOQLEEEAAK	1290	GPNTAPPNFSTGTPTFPVVP-----FLSSIAGVPTTAAATAPVATSPSP	1335
DB	99 PLSLQFSPSIPAAVPSQAPPEPREEV-----TATATSQVAAQPPAAAAPGEQAVA	1528	-----VLVPELKGKQPSPL--TYEDHGAPFAGHLPRGSPVTMREPTRELQGSLSSSKA	1579
QY	207 PPEP-----EKVPSPPIESKHSRLVQIYDENRKKAAHRLIEGLGPQVELPLY	1336	NDISTSVIQSEVTVTEEGIAGVATSTGVTSGGI-----PIPVSESPVLSVV	1385
DB	150 GPAPSTVPSSTKDRVSPSL-----VGSKEEPPPA	1580	SQDRKLTSTPREIA---KSPHSTVPEHHPHIPISPYEHLRLRGVGDVLYRSHIPLAFDPTS	1636
QY	258 NPSTDQVHENIKINOARKKLILYFKRNHARKQKQFCQRYDQDLMEALEKKVERIE	1386	SS-----IITPAVVSISTTSPSLQVPTSTSEIV-----VSTALYPS-----	1422
DB	182 RSGSG-----GSAKEPQBERSQQQDDI--EELETKAVGMS	1637	IPRGIPLDAAAAYLPRHLAPNPTPHLYPPYLI-----RGYDDT	1676
QY	318 NNPRRRAKESKV-REYEEKQFPEIRKQ-----ELQRMQSRVQSGSLSMSAARSE	1423	-----YTVSATSASAGGSTATPGPK-----PPAVVSQAGSTTVGATLTSTVTTSTFPST	1473
DB	216 NDGRFLKFDIEICRGFSFKTVYKGLDTFTTVEVAVWCELQDRK-----LTKSERQRF	1677	AALNRQ-----TIINDYITSQQHHNTATAMA-----ORA	1707
QY	370 HEVSEIIDGLSQENLEKQMLQVAVIPMLYDADQORIK-----FINNGLM-ADPMKVYK	1474	ASQLSIQLSSSTSTPTLAEITVVVSAHSLDKTSHSTTGLTGLAFSLAPSSSSSPGAGVSSY	1533
DB	266 KEEAEMKGL-QHPNIVR-----FYDSWESTVAGKKCVILVTELMTSGTLKTYL	1708	DMLRGLSPR-----ESSIALNAAAGPRGIIDLSQVPHLPVLVPPPTGPTATAMDRLAYL	1761
QY	425 DR-----OVNMWMSQEKETREKFMQHPK-----NFGILTA			
DB	314 KRFKWKIKVLRSCWCKQILKGLQFLHTRTPPIIHRDLKCDNIFITGPTGSVKIGDLGL--			
QY	456 SFLEKRTVAECVL-----YYILTQKNENY-----KSLVRSYRRRGKSQQOQQQQQ			
DB	372 ATLKRASFVAKSVIGTPEFWAPMEYEEKYDESVDVYAFGCMLEMATSEVPYSECQNAQI			
QY	502 QOQOQOQOQO-----MPRSSQ-----EKDEKEKEKEAEKKEEKEPEVNDKEDLLK			
DB	432 YRRVTSGVPASFDKVAIPEVKEIIEGCRQNKDERYSIKOLLNHAFFQEBETGVRVELAE			
QY	548 EXTDDTSGEDNDEKAVASGRKTANSQGRKRGRITRSMANEANSE-EALITPQQOAEALAS			
DB	492 E-----DQGEKIAIKMLRIEDIKKLGKGYKQNEAIEFSDFLERDVPEDVAQ----			
QY	607 MELNBSRWTEEBEMETAKGLLEHGRNWSAIRMVGSKTVSQCKNFYFNKQNLNDEIL			
DB	539 -EWESGYCEGDHKTMAIKDR-----VSLIK-----RKREQQLVR			
QY	667 QQHKLKMEKERNARKKKA-----PAAASEEAAFPVVEDEEWEASG			
DB	577 EEQKKKQOESSLKQOQVEQSASQTGKQLPQASSTGIPTASTTSASVSTQVEPEPEEA--			
QY	710 VSGNEEMVEEAELHASGNEVPRGCSGPATVN--NSSDTEIIPSPHTEAAKDTQNGP			
DB	635 -DQHQOLQYQQPSISVLSDGTVDGQGSVFTESSVSSQOQTSYGSQH--EQAHSSTGTVP			

Db 1534 SQPGLHLPLVPSVIASTPILPQAAGSTPTLLPQVPSIPPLVQPVANPAV-----QQTLLI 1590
QY 1762 PTAPQPSRSHSSPLSPGP-THLTKPTTSSSERDRDRDREREKSIILTSTTT 1820
Db 1591 HSQOPQ-----ALLPNQPTHCP-----EVDSTQPKAFIDDIKT 1626
QY 1821 VEHAPIWRGTQSSGSSGSSGSSGSSSRPASHSHAHQSPISPTQDALOQRPVSLHN 1880
Db 1627 LEE-----KLRSLPSEHSSGA-----QHASVSLTS-----LVIES 1658
QY 1881 TGMKGI-ITAVEPSKTVLRSSTSPVRPAATFPATHCPGLGTLGTVPLMEVPLP 1939
Db 1659 TVTPGIPPTAVAPSK-LLTSTT-----SCLPPTNLPLG----- 1691
QY 1940 KEAPRVARERADTGHAFLAKPPARSGLEPASPSKSGSEPRPLVPPVSGHATARTPA 1999
Db 1692 ----TVALPVTVPVVGQVSTPVTSTTSGVKGTAFSKPLKAPVLVGTETLPGTLD 1747
QY 2000 KNLAPHASPDPPAPPASADPHREKTQS-KPFSIQEELRLSLGYHSGSYSPGVPVGP 2058
Db 1748 EQL-----PPFPGPSL-----TQSQPLEDLDAQLR-----RTLSPXITVISA 1786
QY 2059 VSPSLTHDKGLPKHLEELDKSHLEGELRPQKQPVKLGGEAAHLPHLPLPE---SQPS 2115
Db 1787 V-----GPVSMAPTA-ITEAGTQPKGVSVQK 1813
QY 2116 SGPLLOTAPCV-----KGHORVTLAQHISEVITQDTRHHPOQLSAPLPAPLYSPGASC 2171
Db 1814 EGPVLATSSGAGVFKMGFRQVSVAA-----DGAQEGKNSDEKSVFESSTSES 1864
QY 2172 PVLDLRRPESDLYLPP-----DHGAPARGSPHSEGGKRSPEPNKTS----- 2213
Db 1865 SVLSSSSPESTLVKPEPENGITIPGISSDVPESAHTTASEAKSDTGQTKVGRFQVTTTA 1924
QY 2214 -----VLGGEGDIE-----PVSPEGMTEPCHRSVAVPLYRDGQTEPFRMGSKS 2261
Db 1925 NKVGRFSVSKTEDKITDTTKEGVPASFPFMDLEQAVLPAVPIPKKEP-ELSEPSHLN--- 1980
QY 2262 PGNTSOP-PAFTSKLTESNAVKSKOEINKKLNTNREPNYISQPGTEIFNMPALT 2320
Db 1981 -GPSSDPEAFUSROVDGSGSPHSQJLSSKSL-----PQNSUSQLSNSFNFSYMS 2032
QY 2321 GTGLMYSRQAVQ-----EHASTNMGLEAIIRKALMGKTDQWESPPLSANAFNPLN 2372
Db 2033 SDNESDIEDEDLKLELRRLRDKHLKEIQDLQSRQKHEIESLYTKLGKVP----- 2082
QY 2373 ASASLPAAMPITADORSHTLTSPOGGGKAKVSGRPSSRKAKSP-APGLASGDRPPSV- 2430
Db 2083 -AVIIPPAAPLS---GRRRRPTKS-----KGSKSRSSSLGNKNSPOLSGNLQCSAASVL 2133
QY 2431 ---SSVHSEGD 2438
Db 2134 HFQQTLLHPGN 2144
RESULT 80
ABB50164
ID ABB50164 standard; protein; 1828 AA.
XX
AC ABB50164;
XX
DT 05-FEB-2002 (first entry)
DE Human transcription factor TRFX-15.
DE Human; transcription factor; TRFX; cell proliferative disease;
KW autoimmune disease; inflammation; neurological disease;
KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
XX neuroprotective; antiinflammatory; gene therapy.
OS Homo sapiens.
XX

PN WO200172777-A2.
XX
PD 04-OCT-2001.
XX
XX 13-MAR-2001; 2001WO-US008117.
XX
XX 13-MAR-2000; 2000US-0189986P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JL, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;
PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;
PI Reddy R;
XX
DR WPI; 2001-570896/64.
DR N-PSDB; ABA82988.
XX
PT Novel transcription factor polypeptides, used to treat diseases
PT associated with altered activity and expression of TRFX, and to screen
PT for agents capable of modulating its activity.
XX
PS Claim 1; Page 154-158; 327pp; English.
XX
CC The present sequence is the protein sequence for a human transcription
CC factor. The transcription factor and its coding sequence are useful in
CC the diagnosis, treatment and prevention of diseases associated with
CC altered expression of the transcription factor e.g. cell proliferative,
CC autoimmune/inflammatory, neurological and developmental disorders. A
CC number of specific disorders/diseases are given in the specification,
CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
CC disease, stroke, and viral, bacterial, fungal and protozoal infections
XX
SQ Sequence 1828 AA;

Query Match 3.2%; Score 418; DB 4; Length 1828;
Best Local Similarity 19.8%; Pred. No. 2.4e-14;
Matches 402; Conservative 211; Mismatches 719; Indels 694; Gaps 80;
QY 687 PAAAEAEAFPPVVEDEMEASGVSGNEEEMVEAEALHASGVNEVPRGCSGATVNNSS 746
Db 40 PSSLPGPASSPWPINSSPLASPVSTVS--VPLSSSLPISVPTTLPAPASAPLTI---- 93
QY 747 DTESIPSPITEAKD-----TGQNGPKPPATLGADGPPPGPPT-----PP 786
Db 94 ---PISAPLTVSASGALLTSTVTPPLAPVVPVPAAPGPPSLAPSGASPSASALTGLATAPS 150
QY 787 RRTSRAPIEP-----TPASEATGAPTPP-----APPSAP---PPVVPKEE 826
Db 151 LSSSTQPGHLLAPTSSTSHVPGNLSTVAPACGVLVPASALASPPFAPNPAFLPLPV 210
QY 827 KEEETAAPVVEEGEQKPPAAEELAVDTGKAE-EPVKSECTEEAEAG-----PAK 876
Db 211 LAPSGAAPVLASSQTFVPVMAPSSTPGTSLASASVPVAPTPVLPASSTQTMPLPAPVSP 270
QY 877 GKDAEAAEATAGALKAEKKEGSGRATTAKSGCAPQDSSSATSDEVEDEAGDKNR 936
Db 271 LPSPASTQTLALAPALAPTLGGSSPSQTLSLGTGNPQGPPTQTLS----- 316
QY 937 LLSRPSLLTPTGDPANASPOKPLDLKOLKQRAAAIPIQVTKVHPREDAAPTAP 996
Db 317 -ITPSSLV-PTPAQTLSLAPGPPGLPTQTLSLAPA-PPL-----APASVPG 360
QY 997 PAPPFPQNLPQ-ESDAPQPGSSPRGKSRSPAP-----PADKEAFAAEAQ 1040
Db 361 PAPAHTLTLAPASSASLLAPASVQTLTSLPAPVPTLGPAAQTALAPASTQSPASQAS 420
QY 1041 KLPQDPPCWTSGLPFPV-----PPREVIKASHAPDPSAFSA----- 1078

CC	aureus. (I) are more hydrophilic than known avilamycins. The present
CC	sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC	viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ237515-
CC	ABZ37516)
XX	
SQ	Sequence 19938 AA;
	Query Match 3.2%; Score 417.5; DB 6; Length 19938;
	Best Local Similarity 20.5%; Pred. No. 5.9e-13;
	Matches 476; Conservative 158; Mismatches 813; Indels 879; Gaps 114;
QY	603 ELASMELNNESSRTEEREMETAKKGLLEHGRNWSAIRMVSGKTVSQCNFYFNKKRONL 662
DB	1420 ELGQHEIQEAAA--EEEPVDHADQGQADAGRPQQGGHLA-----1456
QY	663 DEILQOHLKWEKERNARRKKXAPAAASEEAAPPVVVEDEMEASGVSGNEEBEMEVAE 722
DB	1457 -EPFGQSPRRGDHRVDRTGEAAPARTPGAGTRPHGHV---GCRGR-----1504
QY	723 ALHASGNVP--RGECSG-----PATVNNSDITESIPSPHTE-----757
DB	1505 --HAGDRPVPAVGAVHGARAGGHL PVLVRQRPARPAPDLRLPRGPARGARTAVRRGRA 1562
QY	758 -----AAKDTGQNGKPAPA-----TLGADGPPGPTPPPRRTSRAPIEPT 797
DB	1563 GSODGGSHRTFGDPAPQOVVADDELHYGCAGAVHRTGAES--PPRAPPPPFRVRTPGARP 1621
QY	798 PAS-----EATGAPTPPPAPSPSAPPVVPVKKEEETAAPPEEGEEKPKPA---847
DB	1622 PVSTHAHLRTRTGRPLHPG--DARSFAPGLHREPTQVPWFV----GRGVHPGAQVC 1675
QY	848 -ABELAVDTGKAEEPVK-----SECTEEA-----870
DB	1676 DAAVTGHDISAGVEEPARPSGRXWMTALEWFKSSYSGSEGQCQVEVALCPHTTIHSDKN 1735
QY	871 --REGPAKGKDAA-----ASATBAGALKAE-----KKEGSGR-----ATTA 906
DB	1736 TPEDGPTLOVSPATAWRAFTSATTEXARRBELTPTVTPVPPPCGGCGRCRRRPVSTR 1795
QY	907 KSSCAPQDSSTATCSADEVDAE-----GGDKN-----935
DB	1796 RPCGGKRKTGTGTCSPASAPGVVPPPLPCRAGGARGSPGCGGHRPGVRRSAXAAA 1855
QY	936 RLISP--RPSLLTPTGDPANASPKQLDLKLQKQRAA-----AIP--974
DB	1856 RGSPRRPACXGPLRLPRSARAVRRDRARRSERRAAVGGRAACRPCRSARTHARPAVPA 1915
QY	975 -----PIQVTKVHEPREDAATKPA-----PPAP-----P 1000
DB	1916 PCAAPSAPSAFCVCFVRRSP-----GPACPAWLFPFPFGPAHPLWLALRTPARPAGLLP 1971
QY	1001 P-----PQNLPQESDAPQPGS--SPRGKSRSAPPADKEAFAAEQALPG--DPCWTSLG 1053
DB	1972 PLARPRFXARVFLVFVPGTGTRTPRARSRRPPQRSRAAGAPSTADPARTPPC-----2026
QY	1054 PFVPVPREVIKASHAPADPSAFSYAPP-----GHPPLGLHDTARPVL-----1096
DB	2027 ---ASPASCRRARPRGRPPRPPCWAPTRPCAGAPGVXRGPAGCGADRVPVPCCGTPWS 2083
QY	1097 -PRPTTSNPPLLISSAKHSVLERQTGAIISQMSVOLHVPY---SEHAKAPGVPTVMGL 1152
DB	2084 PPRRRARRWPFPARAGSAPG-----CCTPRRGS-----PWYRSGTARGRSAPPTRRR 2132
QY	1153 PL-----PMDPKKLAPFSGVKQEQQLSPRCQAQPESIGVPTAQ-EA 1192
DB	2133 PWMGGSGDSRSRGRPPRRACPFRPRR--APGPSAARRRWPRRRPSSRRTRPAGRGRA 2190
QY	1193 SVLRGTALGS-VPGGSIITKGIPSTRVPSDAITYRGSITHGTADLVLYKGTITRIIGEDS 1251
DB	2191 AVSRAAHCAERSFGSSXSARWFDPPXSGSAPRELGARVRWP-----GSSCR-----A 2239
QY	1252 PSRLDRGREDSLPKGHVIFYGKKGHVLUSYGGMSVT---QCSKEDGRSSSGPPHETAAPK 1308

Db 733 PGSPHMAQOH-----GDPAITANDVLSQMPPDVSIQOTNNVPPHVQAMQ 780
QY 1539 ---SPLTYEDHG---APFAG-----HLPRGSPVMTREB-TPRLQEGSLSS 1577
Db 781 NSASGNHFSHGMSFNAPSGAPNGQMGQNGPFFVNVKDVTLTSPLLVNLQSDISAG 840
QY 1578 KASQDRKLTST-----PREIAKSPHS-TVPEHHPHIPISYEH-LLRGVSGVDLYR 1625
Db 841 HFGVNNKQNTWANKPKKPKKNSQODLNTPTDTRPAGLEAEOPLPGEQGISLDN 900
QY 1626 SHIPLAFDPTSI PRGI---PLDAAAAYLP---RHLPANPTVPHLYP-PYLIR----- 1671
Db 901 SG-PKLPFSPNPPGVPSPQVEPRLQMPFPQMLMQHVAAPPPOPPQOQPOLPQOQPPP 959
QY 1672 -GYPDTAALENROTIINDYITTSQMHNTATAMAQADMLRG-LSPRESSLALNVAAGPR 1729
Db 960 PSQPSQOQQOQQOQQMMMLMQDPKSVRLPVSNVHPHPPRGPLNDSQRMPPQSGSV 1019
QY 1730 GIIDLQVPHLPVLVPPTP-----GTPATAMDRLAYLPTAQO-PFSS----- 1770
Db 1020 VMVSLQG-----PASVPPSPDKQMPMPVNTPLGNSRKMVYQSPQNPSSSPLAEMASLP 1075
QY 1771 --RHSSSPLSPGCPHTLTKPTTTSSEERDRDRDREREKSLTSTTVEHAPIWR 1828
Db 1076 EASGSEAPSVPGSPNNMPSHV-----VLPQNQLMMTGP--K 1109
QY 1829 PGTEQSSGSGSGSGSGSGSPASHAHQSP-ISPRTOALQORPSVLHNTGMKGII 1887
Db 1110 PGSPLSATQCATPQPPVNSLPSSGH--HFNVAAPTQTS---RPKTPNRPSPRY 1163
QY 1888 TAVPSKPTVLRSTSTS-SFVRPAATPTPATHCFLGGTLGVVYPTLMBPVLLPKEAPRVA 1946
Db 1164 PQTNNRPPSPSPSLSPER-----LNASTAGLFP-----POIN 1199
QY 1947 RPERPADTGHAFLAPPARSGLEPASPSKSGSEPRPLVPVSGHATIAITPAKNLAPH 2006
Db 1200 IELPPRPNLRGP-----DOQGLNPTTLKAIGQAPNSLNTWNSNFAT----- 1241
QY 2007 ASPDPAPPASADPHREKTQSKPFQIELESLGYHSGSVSPGVEPVPSPVSPSLTH 2066
Db 1242 -----PQTHKLDVVVN-----SKQNSGATKRAFSNS----- 1271
QY 2067 DKGLPKHLEELKSHLEGELRPQPGVKLGGEAAHLPHLRPLPESQPSSSPLLTAPGV 2126
Db 1272 -----RRSSPGSR-----KTTPSPGRQNSKAPKL----- 1296
QY 2127 KGHQVVTLAHLSEVITQDYTRHHQQLSAPLPAFLYPPGACFPVLDLRPPSDLYLP 2186
Db 1297 -----TLASQTNALLO-----NVELPRNVLVSP 1320
QY 2187 PPDHGAPARGS-PHSEGGKRSPPNKTSVLGGG---EDGIEPVSP----- 2228
Db 1321 TPLANFPVPGSPFNNSG--LNFONSTVVAAGVVEDNKESLNVQDSDCQNSQSRKEQ 1378
QY 2229 ---EGMTEPGHRSAYVP--LLYRDGEQTEPSRMGSKSPGNTSQPPAPFSKLTESAMV 2283
Db 1379 VNIELKAVPAQEVKVMVPEQSKQDQSPDNKLP-----VEENKLV 1422
QY 2284 KSKQKQINKLN--THNREPEYNIQPG-TEI-FNMPALTGTGLM-----TYRSQAVQ 2333
Db 1423 SPAMREAPTSLSQLDNGSAPNTIKPPGLTDLVTPPVVSGEDLKKASVIPTLQDLSS 1482
QY 2334 EHASTNMGIL-----EATIRKALMGKYDQWES-----PPLSANAF 2368
Db 1483 KEPSNGLNPHSNELCSLVHPSESVNSNVAFSIPVMWRPVSSSSISTPLPPNQITVF 1542
QY 2369 ---NPLNASASLPAAMPITAAADGRSDHLLTSPGGGKAKVS-----GRP----- 2409
Db 1543 VTSNPITTSANTSAAALPHTLSALMSTVWTPMAGSKVMVSEQSAQAQSNARQFITTPE 1602
QY 2410 -----SRKAKSPAPGLA--SGDRPPSVSV--HSEGDGN-RRTPLTNRWED 2452
Db 1603 INSSIIQVMKGSQSPSTIPAAPLTTNSGLMPPSVAVVGLHIPQNIKFSSAPV----- 1656

QY 2453 RPSSAGSTFFPY-----NPLIMELQAGVMASPPPPGLPAGSGPLAGHAWDEBPKPL 2505
Db 1657 -ENALSSSPAPNIQTGRPLVLSRATPVQLPSP---PCTSSPVV-PSHPVPVQOVKEL 1708
RESULT 83
ABO52986
ID ABO52986 standard; protein; 1791 AA.
XX ABO52986;
AC ABO52986;
XX 09-OCT-2003 (first entry)
XX Human putative spliceosome associated protein (SAP) #48.
KW Human; SAP; spliceosome associated protein; ribonucleoprotein;
KW RNP complex; RNA affinity substrate; RNP assembly sequence;
KW spliceosomal complex; hnRNP complex; mRNA export complex;
KW mRNA localisation complex; RNA editing complex; intron complex;
KW H complex; telomerase complex; fragile X protein complex;
KW reverse transcriptase complex; gene splicing complex.
XX Homo sapiens.
OS US2003068803-A1.
XX 10-APR-2003.
XX 14-JAN-2002; 2002US-00047991.
XX 12-JAN-2001; 2001US-0261521P.
XX (REED/) REED R.
XX (ZHOU/) ZHOU Z.
XX Reed R, Zhou Z;
XX WPI; 2003-540885/51.
XX Isolating ribonucleoprotein complex, by contacting RNA affinity substrate having ribonucleoprotein assembly sequence and affinity tag, with protein mixture, subjecting complex formed to chromatography, affinity selection.
XX Claim 24; Page; 39pp; English.
CC The invention relates to forming (M1) an isolated ribonucleoprotein (RNP) complex (C), involves contacting an RNA affinity substrate (S) comprising an RNP assembly sequence (AS) and an affinity tag, with a protein mixture to permit formation of (C) on AS, subjecting (C) to chromatographic separation, and subjecting (C) to affinity selection, where the affinity tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli maltose binding protein) binds to an affinity matrix. Also included are an isolated spliceosome preparation (isolated by (M1)), a RNA comprising an RNP complex binding site and at least one phage coat protein recognition site, a nucleic acid encoding the RNA, and treating (M2) a subject having a disorder associated with abnormal RNP complexes (by obtaining a sample of cells from a subject, purifying RNP complexes from the cells of the subject by (M1), determining the presence in the purified RNP complexes of one or more proteins, and normalising the amount of RNPs in the subject. (M1) is useful for forming an isolated RNP complex selected from a spliceosomal complex (selected from E, A, B and C complex), an hnRNP complex, an mRNA export complex, an mRNA localisation complex, an RNA editing complex, an intron complex, or an H complex. (M1) is useful in a diagnostic assay for determining whether a subject has a abnormal RNP complexes, (M2) is useful for treating a subject having a disorder associated with abnormal RNP complexes. (M1) is useful for forming an isolated RNP complex such as a telomerase complex, a fragile X protein complex, a reverse transcriptase complex or a gene splicing complex. The present sequence represents a putative novel human spliceosome associated protein (SAP) isolated by the methods of the invention. Note: The present sequence is not shown in the specification but was obtained from Genbank or Swissprot using the information provided

QY 1639 RGIPLDAAAAYLPHRLANPTYPHYLPYLRGYPDTAALNKRQTIINDYITSQOMHIN 1698
Db 1146 QPQP-----PPPPPPQPALPS--PPPLVAPT----- 1171
QY 1699 TATAMAQRADMLRGLSPRESSLALNYAAGPRGIIIDLSQVPHLPVLVPPTPGTATAMDR 1758
Db 1172 -----SSPPPPPLPPPPPPAMPSP----- 1190
QY 1759 AYLPTAQPFSSRHSSPLSPGQPT-----HLTKPTTSSSRERDRD----- 1801
Db 1191 -----PPPPPPAAAPLAAPPEPAAPSPEDPELPTRELHLAKQOETAACVGETDEEAGSG 1247
QY 1802 -----RERDREREKSI-----LTSTTVEHAPIWR-----PGTEQSSGSS 1838
Db 1248 GEGIFRERDEFVIRAEIDPSLKALQGTGREPPPIWRVQKALLQKFTPEIKDQORQFCATS 1307
QY 1839 GSSGGGGSSSPASHAHO-----HSPISPRQTQDALQOR 1874
Db 1308 NYLGVFGDAKNR-----YQRLYVKFLENVKKOYVRVCARKPHRPPVVRSGOAKN 1360
QY 1875 PSLVHTMGKGIITAVEPSKP--TVLRSTSTSPVRPAATFPFATHCPLGGTLDGVVPTLM 1933
Db 1361 PVSAGSSAPPKAPAPPKPEPTTSEKPEQTETAMPEPPAP-----EKPSLL 1413
QY 1934 EPVLLPKEAPRVARPERP-----RADTG-----HAFKAPPARSGLEPASPSPK-GSEPR 1982
Db 1414 RPVEKEKEKVTGRERPLRGERATSGRQTPERSLATQGPATSL- PKARPTKVKAEPP 1472
QY 1983 P-----LVPVVSCHATIATPAKNLAPHASDPDPAPPASADPHREKTQS----- 2028
Db 1473 PKRKKWLKEAGNATAGGPGSSSDSSPGAPSEDERAVFGRLLKTRAMRMYRSV 1532
QY 2029 -----KPFSTQLELESLGYSYSPGVEPVSPVSPSLTHDKGLPKHLE-----E 2076
Db 1533 EMLVSTALPDMDIALE-----DTHDELYP-----PMRKIDGLLNEHKVKLXLSLSPAL 1584
QY 2077 LDKSHLEGLRPKQPGVKLGEEAAHLPHRLP--PESQPSPLLOTAPGVKGHORVVT 2134
Db 1585 QDALHTFPQLQVEQSGE--GSPEGAVRLRPAEPYNRKTLKL-----KRSVV 1631
QY 2135 LAQHI 2139
Db 1632 RAQEV 1636

RESULT 86
ABM78990
ID ABM78990 standard; protein; 2092 AA.
AC ABM78990;
DT 15-JAN-2004 (first entry)
TX
DE Modifier of beta-catenin (MBCAT) polypeptide GI 27498257.
KW Modifier of beta-catenin; MBCAT; protein kinase; enzyme; cytostatic.
XX Homo sapiens.
XX
FH Key
FT Domain
FT 39..312
FT /note= "kinase domain"
XX
PN WO2003074725-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006294.
XX
PR 01-MAR-2002; 2002US-0361242P.
XX
PA (EXEL-) EXBLIXIS INC.
XX

PI Costa MA, Gendreau SB, Dora EG, Nicoll M, Heuer TS;
DR WPI; 2003-756771/71.
XX
XX Identifying a candidate beta-catenin pathway-modulating agent, for
PT diagnosing and treating cancer, comprises contacting an assay system
PT comprising a modifier of beta-catenin (MBCAT) polypeptide or nucleic acid
PT with a test agent.
XX
XX Claim 13; Page 105-114; 114pp; English.
XX
XX The present sequence is that of human modifier of beta-catenin (MBCAT)
CC polypeptide GI 27498257. MBCATs are kinase proteins with protein kinase
CC domains. Genetic screens were designed to identify modifiers of the beta-
CC catenin pathway in *Caenorhabditis elegans*, and the C10C6.1 gene was
CC identified. Human orthologues, including the present sequence, were
CC subsequently obtained. MBCAT genes and polypeptides can be used to
CC identify MBCAT-modulating agents that are candidate therapeutic agents
CC for treatment of disorders associated with defective or impaired beta-
CC catenin and/or MBCAT function, such as an angiogenic, apoptotic or cell
CC proliferation disorder, e.g. breast, colon, head and neck, kidney, lung,
CC ovarian, prostate, skin or uterine cancer. MBCAT modulating agents
CC include antisense oligomers and RNAi that repress gene expression or
CC product activity. Transgenic non-human animals are useful as models of
CC disease and disorders implicating defective beta-catenin function
XX
XX Sequence 2092 AA;
XX
Query Match 3.1%; Score 414; DB 7; Length 2092;
Best Local Similarity 20.0%; Pred. No. 4,8e-14;
Matches 452; Conservative 245; Mismatches 787; Indels 780; Gaps 112;
QY 471 YLTKNENKSVLRYSRRRKSQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 525
Db 304 YEVKQHRFPRSLDWSLLR-----QKAEFIPQLESEDDTSYFDRSEK 346
QY 526 EXEAKSEBEKPEVNDKEDLLKEKTDGSDG-NDKEAVAS---KGRKTANSQGRKGR 581
Db 347 YHHMETEEB-----DDTNDEDFNVEIROFSCSHRFKVFSS---IDR 386
QY 582 ITRSMANEANSEAITPOQSAELASWE-LNESSRWTE-EEMETAKKGLLEHGNWSAIAR 639
Db 387 ITQNSAAE--KEDSVDKTKSTLPSTETLSWSSEYSEMOQLSTNSDSTESNR----- 437
QY 640 MVGSKTVSOCKNFYNYKKRQNLDEILOQHLKMKERNARKKKKAPAAAEEAAPPV 699
Db 438 -----HKLSSGL---LPKLAISTEGEQD---EAASCQDPHEEPCKPAL 475
QY 700 VEDEMEASGVSGNBEEMVVEABALHASGNEV-----PRGCSGPATVNNSSDT 748
Db 476 PPEE-----CAQEBPEVTTASTTSSSTLSVGSFSEHLQDINGRSECDV--STDNSSKP 527
QY 749 ESIPSPH-----TEAKDTGQNGKPPATLGADG-----PPGPPPTPRR 788
Db 528 SSEPASHMARQLESTKKGISG---KVTLSLSASALSMIPGDMFAVSPGLGSPMSPHS 583
QY 789 TSRAPIEPTPASEATCAPTPPPAPPSPSPAPPVVPKKEKEE--TAAAPVVEGE----- 841
Db 584 LS-----SDPSSSRDSSPSRDSASASAPHQIVIHSSCKNYGFTIRALVYVGDSDIYT 638
QY 842 -----EQKPPAAE-----ELAVDTGK-----ABEPVK 863
Db 639 VHHIVMNVBEGSPACQAGLKAGDLITHINGEPVHGLVHTEVIELLLKSGNKVSIITTPPE 698
QY 864 SECTEAEEGPAKGDAAEAAEAETACALKAEEKGGSGRAT-----TAKSSGAPQDSSSA 919
Db 699 N---TSIKTGPARNSYKS--RMVRRSKKSKKESLERRRSLFKKLAKQPSPLHTSRSP 753
QY 920 TCSADEVDEAGGDKNRLLSPRPSLLTPTGDPANASPOKPLDLKQLKQRAAAPIPIQVT 979
Db 754 SC-----LNRSLSGSEL---PGSTHLSLPRSP----- 779
QY 980 KVHEPPREDAPTKPAPPAPPPQNLQPEDAPQQFGSSPRGKSRSP-----APPADKEA 1034

Db 780 ----TFSYRTPDFPS--GTNSQSSSPSSAANS--GHRSTLHGLAPKLGQR 831
QY 1035 FAABAQKLPDPPCWITGSLFPVPPREVIVKASPHA--PPPSAFSYAPPCHPIPLGLHDTA 1092
Db 832 YRSGRKRSAGNIP-----LSPLARTPPTPOPTSPQRSPLLGHSLG 874
QY 1093 RPYLPR--PTTISNPPLISSAKHPSVLEROIGAIQSGMSVOLHVPYSEHAKAPVCPVTM 1150
Db 875 NSKIAQAFPSKMSPTTIVR-----HIVRPKSABPPRSPL-- 909
QY 1151 GLPLPMDPKKLAPFSGVKBQOLSPRQAGPPESLGVPTAQEASVLRTALGSPVGGSIK 1210
Db 910 -LKRVOSEELKSPSYGDKKHLCSR-----KHSLEV-TQEE--VQREQOREAPLOSIDE 960
QY 1211 GIPSTRVPSDAITYGSGITHGTPADVLYKGTITRIIG--EDGPSRLDRGREDSLPKGHVI 1269
Db 961 NVCDFPPLSRAREVEQGCL-----KRPVSRKVGRQESVDDILDR--DKL-KAKVV 1006
QY 1270 YEGKGHVLSEYEGMSVTQCKEDGRSSSGPPHETAAPKRTYDMMGRVGRASSASIEG 1329
Db 1007 VK-----KADGF----- 1013
QY 1330 LMGRAIPPRHSHPHLKEOHIRGISTIGIPRSYEAQEDYLRRKALLKREGTTPPPPP 1389
Db 1014 -----PEKQESH--QKSHGPGSDLENFALFKLE-----EREKKVY-----PK 1048
QY 1390 SRDLTAYKTOA-----LGPLKPAHEGLVATVKEAGRSIHEIPRELHRTPELPIA 1442
Db 1049 AVERSSTFENKASQMBAPPLGSLLDKALHQAQSVASEGAMSDGVPFAEHRQGGGDFRA 1108
QY 1443 PRELKEGSIQTGPLKYDVTGASTTG--SKKHVRSVLIGSPGRTTPPVHPLDVMADARAL 1499
Db 1109 PAP--GTLQDGLCHSLDRGISGKBGTEKSSQAKELLRCE--KLDSKLANIDYLRKQMSL 1164
QY 1500 ERACYEESLKSRPGTASSGGSIGARGPVIVPELKGPROSLPTYEDHGAPFAFAGHLPGRSP 1559
Db 1165 EDK--EDNL-----CPVL-----KPKMTAGSHE-----CLP-GNP 1191
QY 1560 VT-----MREPTPLQEGS--LSSSKASQDRKLSTPREI-----AKSPHSTVVEHHPH 1606
Db 1192 VRPTGQQQPPPPASERAFVSSTHAQAQMSAVFVPLKALTGRVDSGTEKPGGLVAPE-SPV 1250
QY 1607 PISPVHEHLRG-----VSGVDLVRSHP--LAFDPT-----SIPR 1639
Db 1251 RKSPEYKLEGRSVCLKPIEGTLDIALSGQASKTELPSPESAQSPSGDVASVPP 1310
QY 1640 GIPL-----DAAAYYL--PRHAPNPTYPHLYPPYIRGYPDTAALENRQTII---NDY 1689
Db 1311 VLPSSGKKNDDTSARELSFSSLKMNKSY--LLEPWFL---PPSRGLQNSPAVSLPDPEF 1365
QY 1690 ITSQOMHNTA---TAMQAQADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLV 1745
Db 1366 KDRKGPHPTARSPTGVMESENPOQREGSGPKHQ-----DHTTDPKLLTCLGQNLHSPDLAR 1421
QY 1746 PTFGTATAMDLAVLPTAPQPFSSRHSSSPLSPGPGTHLTKPTTTSSSER-----ER 1798
Db 1422 P-----RCPLPPEASPSREKPGLRRESSERGPTARGER 1454
QY 1799 DRDR-----ERDRDREREKIL-----TSTTTVEHAPITWRPGTQGS 1834
Db 1455 SAARADTCHREPSMELCFPETAKTSDNSKNLLSVGRTHPDFTYTQAMEKA--WAPG---- 1508
QY 1835 SGSSGSSGGGSSRRPASHSHAHQHSPISPRTDALQORPSVLHNTGMKGIITAVEPSK 1894
Db 1509 -GKTHNKGQPG--EARPPPRDMSLSHAGICEKELGKVR-----RQVERPK 1552
QY 1895 PTVLRSTSTSPVRPAATPPATHCFLGTLGDTGVPTLM-----BPVLLPKAEAPVARPE 1949
Db 1553 EALLARRSLQ-----PPGIESEKSEKLSS--PFSLQKDGAKPE--RKEQPLQRHPS 1600
QY 1950 R-PRADTGHAFKAPAR-----SGLEPASSPS---KGSEPRPLVPVPSGHATIAR 1996

Db 1601 SIPPPLPTAKDLSSPAARQHCSSPSHASGREPGAKPSTAEPSPPQDPPKPAVAAHS----- 1656
QY 1997 TPAKNLAPHASDPDPAPASADPHREKTKQSKPFSI-----OELELRSLG---YHGSSY 2048
Db 1657 ----ESSGHKPRPGDPGPKTKHDPDRSLSSQKP--SVGATKGKEPATQSLGSSREGKCH 1711
QY 2049 SPFG--VEPVSPVS---SPSLTHDKGLPKHLEELDLSHLEGELELPRKQPG--PVKLGGEAA 2101
Db 1712 SKSGPDVFPATPGSQNKASDGIQGEQGVSVPLHTDRAPLDAKPOQTSGGREFLEVKPV 1771
QY 2102 HLPHLRPLPESOPSSSPLLQTAPGVKGHQRVVTLLAQHISEVITQDYTRHHPOOLSAPLPA 2161
Db 1772 HLPFRCHGCPSEPAQKL--SAVEK-----QTLSPKHFK-----PS 1806
QY 2162 PLYSPFGASCPLDLR---RPPSDLYLPPDHGAPARGSPHSEGGKGRSPE-----PNKTSV 2214
Db 1807 TVKDCP-TLKQTDNRQTDKSPSQ-----PAANTDRAEGKKCTEALYAPAEQDK 1855
QY 2215 LGGGEDGIEPVSPPEGMPTEPGHSRAVYVLLYRDGEQTEPSRMGSKSQNTSQPPAFPSK 2274
Db 1856 LEAGLSFVHSENRLLKGAERPAAGVGKGF-----EARGKGPQ-PQKPP----- 1897
QY 2275 LTESNSAMVSKKQKQINKKLANTHNREPEYNISQPTGIFNMPAITGTCLMTYRQAOVE 2334
Db 1898 -TEAD-----KPNGMKESP-----SATQSSSFRSTALPE 1925
QY 2335 H-----ASTNMGLEAIRKALMKQYDQWESPPLSANAFNPLNASASLPAAMPITAADGRS 2390
Db 1926 KSLSCSSFPETRAGVREASASSD-----TSSAKAAGGMLLELPAPSNR 1969
QY 2391 DHTLTSPPGGGK--AKVSGRPSR----- 2412
Db 1970 DHRKAQPAGEBRTHTMTKSDSLPSFRVSTLPLESHHPDNTMGGASHRDRALSVTATVGET 2029
QY 2413 KAKSPAPGLASGDRPSSVSVHSEGDNCNRRPTLTNRVWEDRPSS 2456
Db 2030 KGKDPAPA-----QPPARKQNVGRDVTKPSAPN---TDRPIS 2065

RESULT 87
AAM78959
ID AAM78959 standard; protein; 2161 AA.
XX AC AAM78959;
XX DT 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 1621.
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US004098.
XX PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00653561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
PA XX

Db	877	----	TPSYRSTDPFSS--GTNSSQSSPSSSAPNSPAGS--GHIRESTLHGLAPKLGGQR	928
Qy	1035	FAAEAOKLPGDPPCWTSGLFPVPREVIKASPHA--PDPSAFSYAPGCHPLPLGLHDTA	1092	
Db	929	YRSGRRKSAQNP-----LSPLRATPSTPPTQSFQRSPSPLLGHSLG	971	
Qy	1093	RPVLPR--PTTISGNPPPLISSAKHPSVLERIQIGAISQGMVQLHVPIYSEHAKAPVGPVTM	1150	
Db	972	NSKIAQAFPSKMHSPPTIVR-----HIVRPKSABPRPSPL--	1006	
Qy	1151	GLPLPMDPKLIARFSGVGKQEQLSRPGQAGPPESLGVPTAQEAASVLRGTALGSLVPGGSIYK	1210	
Db	1007	--LRVQSEEEKLSPYSQDKKHLCSR-----KHSLV--TOEE--VQREQSOREAPLOSIDE	1057	
Qy	1211	GIPSTRVPSDAITYRGSITHGTIPADVLVYKGTITRIIG--EDPSRLDRGREDLSLPGHVI	1269	
Db	1058	NVCDVPLSPARVPEQCL-----KRPVSRKVGROESVDDLLDR---DKL-KAKVV	1103	
Qy	1270	YEGKKGHVLSYEGGMSVTOCKEDGRSSSGPPHETAAPKRTYDMWEGRVGRAISSASIEG	1329	
Db	1104	VK-----KADGF-----	1110	
Qy	1330	LMGRAIPPERHSPHILKEQHIIIGSITOGIPRSYVEAQEDYLBREAKLLKREGTTPPPPP	1389	
Db	1111	-----PEKQESH--QKPHGFGSDLENFALFKLE-----BREKKVY-----PK	1145	
Qy	1390	SRDLTEAYKTOA-----LGPLKLPAHGLVATVKEAGRSITHEIPRELRHTPELPLA	1442	
Db	1146	AVRSSTFENKASMQEAPPLGSLKDALHKAQASVRASEGAMSGDPVAEHRQGGGDFRRA	1205	
Qy	1443	PRPLKEGSITQGTPLKYDTCASTTG--SKKHQDVRSILIGSGRTFPPVPHLDVMADARAL	1499	
Db	1206	PAP--GTLQDGLCHSLDRGISGKGEGTEKSSQAKELLRCE--KLDKLANIDYLRKKMSL	1261	
Qy	1500	ERACYBESLKSRRPCTASSSGGSIARGAPVIVPELGPXQRQSPFLIYEDHGAPFAGHLPRGSP	1559	
Db	1262	EDK--EDNL-----CPVL-----KPKMTAGSHE-----CLP-GNP	1288	
Qy	1560	VT---WRETPRLQEGS--LSSSKASQDRKLTSTPREI-----AKSPHSTVPEHHPH	1606	
Db	1289	VRPTGGQOEPPPASESRAFVSSTHAAQMSAVFVPLKALTGRVDGSTEKPEGLVAPE-SPV	1347	
Qy	1607	PISPYEHLLRG-----VSGVDLYRSHIP--LAFDPT-----SIPR	1639	
Db	1348	RKSPSEYKLEGRSVCSLEPIEGTLIDIALSGPQASKTELPSPESAQSPSPSGDVRASVPP	1407	
Qy	1640	GIPL-----DAAAAYYL-PRHLAPNTYPHLYPPYILIRGYDPTAALENQTTI---NDY	1689	
Db	1408	VLPSSSGKQNDTTSARLSPSSLKWNKSY--LLEPWFL--PPSRGLQNSPAVSLPDPEF	1462	
Qy	1690	ITSQOMHNTA---TAMARQADMRLRGLSPRESSLALNYAAGRGIIDLSQVPHLPVLVP	1745	
Db	1463	KDRKGHPHTARSPGTWENSNPOREGSPKHQ-----DHTTDPLKLTCLQNLHSPDLAR	1518	
Qy	1746	PTPGTATAMDRLAYLPTAPQPFSSRRSHSSSPLSPGGPTHLTKPTTTTSSSSR-----ER	1798	
Db	1519	P-----RCPLEPSPKPEGLRESSERGPTARSER	1551	
Qy	1799	DRDR-----ERDRDRERKSII-----TSTTTVEHAPIWRPGTEQS	1834	
Db	1552	SARADYTCREPSMELCPFETAKTSDNSKNLLSVGRTHPDPFYTOTQAMEKA--WAPG---	1605	
Qy	1835	SGSSSGSGGGGSSRRASHAHQHSPISPRTOTDALQQRPSVLHNTGMKGIIITAVEPSK	1894	
Db	1606	-GKTNHKDQGP--EARPPRDNSSLHAGIPCCEKELGVR-----RGVEPKP	1649	
Qy	1895	PTVLRSTSTSSVRPAATFFPATHCPLGGTLDGVYPTLM-----EPVLLPKPEAPRVARPE	1949	
Db	1650	EALLARRSLQ-----PPGTESEKSEKLSS--FPSLQDKGAKEPE--RKEQPLQRHPS	1697	
Qy	1950	R-PRADTGHAFKAPPPAR-----SGLEPASSPS---KGSEPRPLVPVPSGHATIAIR	1996	

1698	Db	1698	STPPPLTLAKDLSPPAARQHCSSPSSHASGRBCKAKPSTABPSSSPDPDPKPVAAHS----	1755	
1997	Qy	1997	TPAKNLAPHHAASPDPPAPPASASDPHREKTSQKPFSSI-----QELBLRSLG-----YHGSSY	2048	
1754	Db	1754	-----ESSSHKPRGPDGPPKTKHPDRSLSSQKP-SVGATKGKEPATQSLGSSREGKGH	1808	
2049	Qy	2049	SPG--VEPVSPVS-----SPSLTHDKGLPKHLEELDKSHLEGLRKPQPG-PVKLGEAA	2101	
1809	Db	1809	SKSGPPVFPATPGSQNKASDGIQGGSGGSPVLPHLTDRAFLDAKPOPTSGRPLEVLEKPV	1868	
2102	Qy	2102	HLPHLRLPLPESOPSSPLQLTAPGVKGHQVRVTLAQHISEVITQDVRHHHPQQLSAPLPA	2161	
1869	Db	1869	HLPRLPHGPGSEPADOKL--SAVGEK-----QTLSPKHPK-----PS	1903	
2162	Qy	2162	PLYSFFGASCPVLDLR---RPPSDLYLPPPDHGAPARGSPHSGGKRSP-----PNKTSV	2214	
1904	Db	1904	TVKDCP-TLCKQTDNRQTDKSPSQ-----PAANTDRRAEGKKCTEALYAPAEGDK	1952	
2215	Qy	2215	LGCGEGCIEPVSPPEGWTBFGHRSVAVYLLYRDGEQTSPSRMGSKSPGNTSQPPAFFSK	2274	
1953	Db	1953	LEAGLSFVHSENRKLKAERPAAGVGVGKFP-----EARGKGPG-PQKPP-----	1994	
2275	Qy	2275	LTESNAAWVKSKQKINKLNTNHRNEPEYNIQPGTEIFNMPAITGTGLMTVRSOAVOE	2334	
1995	Db	1995	-TEAD-----KPGMKRGP-----SATGSSFRSTALPE	2022	
2335	Qy	2335	H-----ASTNMGLEAIRKALMGKYDOWEBSPLSANAFNPLNASLPAAMPITAADGRS	2390	
2023	Db	2023	KSLSGSSSSPPTFAGVREASASD-----TSSAKAAGMLELPAPSNR	2066	
2391	Qy	2391	DHTLTSPGGGK---AKVSGRPSR-----	2412	
2067	Db	2067	DHRKAQPAGEGRTHMTKSDSLPSFRVSTLPLESHHPDPNTMGASHRDRALSVTATVGET	2126	
2413	Qy	2413	KAKSPAGLASGDRPSVSVHSEGCNRRTPLTNRVWEDRPSS	2456	
2127	Db	2127	KGKDPA-----OPPPARKQNVGRDVKPSAPN---TDRPIS	2162	
RESULT 89					
AB014747					
ID	AB014747	standard; protein; 4873 AA.			
XX	AC	AB014747;			
XX	AC				
DT	25-AUG-2003	(first entry)			
XX	XX				
DE	Novel human protein #120.				
XX	XX				
KW	Human; NOV; gene therapy; endocrine related disease; diabetes;				
KW	metabolism-related disease; obesity; central nervous system disorder;				
KW	Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;				
KW	schizophrenia; depression; autoimmune disorder; inflammatory disorder;				
KW	psoriasis; allergy; lupus erythematosus; asthma; cancer;				
KW	inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;				
KW	colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;				
KW	prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;				
KW	lung disease; emphysema; obstructive pulmonary disease; haemophilia;				
XX	stroke; infection.				
XX	Homo sapiens.				
XX	OS				
XX	XX				
PN	WO2003023002-A2.				
XX	XX				
PD	20-MAR-2003.				
XX	XX				
XX	09-SEP-2002;	2002WO-US028539.			
XX	XX				
PR	07-SEP-2001;	2001US-0318120P.			
PR	07-SEP-2001;	2001US-0318130P.			
PR	10-SEP-2001;	2001US-0318430P.			
PR	17-SEP-2001;	2001US-0322636P.			
PR	17-SEP-2001;	2001US-0322781P.			

PR 17-SEP-2001; 2001US-0322816P.
 PR 17-SEP-2001; 2001US-0322817P.
 PR 19-SEP-2001; 2001US-0323519P.
 PR 20-SEP-2001; 2001US-0323631P.
 PR 20-SEP-2001; 2001US-0323636P.
 PR 25-SEP-2001; 2001US-0324969P.
 PR 25-SEP-2001; 2001US-0325091P.
 PR 26-SEP-2001; 2001US-0324990P.
 PR 17-APR-2002; 2002US-0373212P.
 PR 06-SEP-2002; 2002US-00236177.
 XX (CURA-) CURAGEN CORP.
 PA
 XX
 XX
 PI Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
 PI Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
 PI Shimkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
 PI Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigaru M, Alsobrook JP;
 PI Lepley DW, Edinger SR, Burgess CE;
 XX
 XX WPI; 2003-313242/30.
 DR N-PSDB; ACD19440.
 XX
 XX
 PT New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX).
 PT and polynucleotides, useful in gene therapy, e.g. for treating or
 PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
 PT stroke or infections.
 XX
 XX Claim 1; Page 340-341; 586pp; English.
 XX
 CC The invention describes a new isolated polypeptide (NOVX). The NOVX
 CC polypeptide, nucleic acid and antibody are useful as therapeutics,
 CC particularly in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
 CC therapy for treating the disease or condition. In particular, the NOVX
 CC polypeptide or polynucleotide is useful for treating endocrine/
 CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
 CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
 CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
 CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
 CC asthma, inflammatory bowel disease, rheumatoid arthritis or
 CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
 CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
 CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
 CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
 CC These are also useful in developing powerful assay system for functional
 CC analysis of various human disorders, as well as in diagnostic
 CC applications, and for monitoring the effects of drugs during clinical
 CC trials. This is the amino acid sequence of a novel human NOV protein
 XX
 SQ Sequence 4873 AA;
 Query Match 3.1%; Score 413; DB 6; Length 4873;
 Best Local Similarity 19.0%; Pred. No. 1.7e-13;
 Matches 477; Conservative 278; Mismatches 894; Indels 868; Gaps 105;
 QY 552 DTSGDNDEKAEVASKGRKTANSQGRKGRITRSMANEANSEAITPQOASLASMELNE 611
 DB 44 DLSQSEERQIAA---VMSRAQGLPKGSVPPAAA-----ESPMSHRKQLDSSHPK 94
 QY 612 SSRWTEEMETAKGLLEH-----GRWSATARMVGSKTVSQCXNFVNYKK 658
 DB 95 QSGRPDPGRPAQPLGSKSRRTDTFRSEQLKGRSPSTI-----SLKE 137
 QY 659 RQNLDEILOQHLKMB-----KERNARKKKKAPAAASEAAPPVVEDMEASGVSGNE 714
 DB 138 SKSRDTLKEHKSSMMPGLSEWA-----LSAVSSVVNKNFNPDLISDSASOBEETK 191
 QY 715 EEWVEAEALHASG-----NEVPRGCSGPATVNNSSDTEISPSHTEAAKDTG 763
 DB 192 KQKVVQKQKPEGIKPLPQQPPKPIKQGGGRDPLQDGTGPKSISSQPEKIKS-- 249
 QY 764 QNGKPPATLGADPPPGPPTPRTSRA-----PIETPT 798

DB 250 ---QPPGT---GKPIQGTOTPTDHAQLPLQDASRPQTKQADIVRGESVKSPLSP 301
 QY 799 ASEATGAPTPPAPSPSNAPPVVVKKEKEEETAAPVVEEGEEOKPPAAEELAVDTGKA 858
 DB 302 SK-----PPIQPTPGKPPAQPFCHESQGPAPPAQPSGLTKP-----LAQOPGTV 349
 QY 859 EEPVKSECTEBAEE---GPAK-----GKDAEAEATAEGALKAEKKEGGSGRATTAKSS 909
 DB 350 KPPVQPPGTTKPPAPLGPAPKPAQOTGSEKPSSE---OPGPKALAAQPPGVOK-TPAQOP 405
 QY 910 GAOQSDSDSATCSADEVDEAEGGDKNRLSPRSLTTPGDPRANASQPKPLDLKQLKOR 969
 DB 406 G-----PAKPTQVGVTPKPLAQOPGLQS 429
 QY 970 AAAP-PIQVTKVHEPPREDAAPTKAP-----PAPPPONLOPESDAP---QPGSS-PRG 1021
 DB 430 PAKAPGPTKTPAQTKPPSQOQSGSTKPPQOPGPAKPSQ---QPGSTKPPSQOQSAKPSA 487
 QY 1022 KSRSPAPADKEAFAAEAQKLPDPPCWTGSLFPVPPPREVIKASPHAPDPDSAFVAPRG 1081
 DB 488 QPSPAKPS-----AQPF---TKPVSTGFGKPLQPTV---SPSAKQ-----PPS 527
 QY 1082 HPLP---LGLHDTARPVLPRPTIS-----NPPLISSAKHPSVLERQ 1121
 DB 528 QGLPKTICPLCNTTELLHVPKCANFNCTECQTTVCSLCGFNPNPHLTEAKEWLCLNCQ 587
 QY 1122 ICASIQGMSVOLHVPYSEHAKAPVPT-----MGLPLM-----D 1157
 DB 588 MKRALGGDLAP---VPSPPQPKLTAPVTTTSAVSKSSPQOQTSPPKDAAPKQDLSKAPE 645
 QY 1158 PKKLAPFGVQEQI---SPRQQA-----GPPESLGVPTAQEAS-----1193
 DB 646 PKPPPL--VKOPTLHGSPSAKAKOPPADSLSKPAPKPEPSVSEQKAPVADDPKQOP 703
 QY 1194 ---VLRGTLGVSFPGSIITKIPSTRVPS-----DSAITYRGSITHG---TPAD 1236
 DB 704 KMVKPTDVLVSSSATTKPIDPSSKVSQSAEKTTPPLKTSAPKSQSPPTGKVTPTFD 763
 QY 1237 ---VLYKCTIIRIIGEDSPSLDRGREDLSLKGHVIYEGKHGVLVSYEGSGSVTQCSKED 1293
 DB 764 SKAIPRPASDSKILSHHPGSPESKQKQVDP-----VOKKEEPKKAQTKMSPKDAPKM 817
 QY 1294 GRSSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIPPRHSPHILKEQHHR- 1352
 DB 818 PKGSPTP---GPRPT-----AGQVTPTPQOSP---KPQEQSR 850
 QY 1353 ---GSTQIGIPRSYVEAQSD-----YLRREAKLLKREGTP-----1384
 DB 851 FSLNLGSIIDA-PKSQPTTPQETVTGKLFPGCASIFSQASNLISTAGQPGHSGSGGAP 909
 QY 1385 -----PPPPSRDLTEAVKTOALGPKLKAHEGLVATVKEAG 1422
 DB 910 MKQAPAPSOPTSQPKSTQOAPPAPAKSI PVKTEKAPAEKLEPKAE-QAPTVRKTE 968
 QY 1423 RSIHEIP---REELRHTPELPLAPRLKEGSIQTQ-PL-----KYDTGASTTSGK 1469
 DB 969 TEKPPPIKDSKSLTAEPQKAVLPKLEKSPKPESTCPLCKTELNIKSGDPNFTNCTEC 1028
 QY 1470 KHDVRSILGSGRTPFPVHPLDV-----MADARALERACYESLSKSRP 1512
 DB 1029 KNQVNLGCFN---PTPLHTEIQEWLCLNCQTORAISGQLGDIRKMPPA---PSGPKASP 1082
 QY 1513 GTASSGSGSIARGAPVIVPELGPQRSPLTVEDGAPFAGHLPKSGSPVTMEPTP-----1567
 DB 1083 MPVPTESSOKTAVPPQVKLVKQEQEVKTEAEKVI-----LEKVKETLSNEKIPDMWTT 1137
 QY 1568 ---RLQEGSLSSKAS-----QDRKLTSTPREIAKS 1595
 DB 1138 DQKQESKLEKOKASALEKQKPLPEEKLIPEEKIRSEKKPLLEKKPTPEDKLLPE 1197
 QY 1596 PHSIVPEHHHPISFYEHLLRGVSGVDLYRSHIPLAFDPTSI PRGILPDAAYVILPHL 1655

Db 1198 AKTSABEEQKH-----DLLKSQVOIAFEK-----LEGRV 1226
QY 1656 APNPTYPHYLPYLYRGPD--TAALENRQTIINDVITSQOMHNTATAMAQRADMLRGL 1713
Db 1227 APKTVOEGKQPTKMEGLSGTQSLPKEDDKTKTKKEQPQPCAKDQEKED-----1281
QY 1714 SPRESSLALNAAAGRGIDL-----SQVPHLPVLVP-----PT 1747
Db 1282 -DKSDTSSSQPKSPQGLSDTYSSDGISSLGELPSLPTDEKILKGLKDSFQSS 1340
QY 1748 PGTP-----ATMDRLAYLPTAQPF-----SRHSSSLSPGCPPTH 1784
Db 1341 PSSPSDLAKLESTVLSILEAQASTLADSEKXTQPHVEPSPEQPKDQEKTSLEIT 1400
QY 1785 LTKPTTSSSERBD--RRDRD--RRREKSLTSTTT-----VEHA 1824
Db 1401 ISSEIKESQERKDFKDSQDIPSSKDHEKSEFVDITTRRPFYDSVESSSENS 1460
QY 1825 PIWRPTEQSSGSGS-----SGGGG-----SSRPASHSH-----AHQH 1860
Db 1461 PVPQRRTTSVSGSSDEVYKQDSQSGSEEDFIRKQIIMSADSDASGSEDDEFIRNQL 1520
QY 1861 SPISPRDQALQRPVSLHNTGKGLITAVEPSKPTVLRSTSTSSVRPAATPPATHCP 1920
Db 1521 KEISSTESQKKE-----ETKGKKITAGKHRLT--RKSTTS-----1556
QY 1921 LGTLDGVPTLMPEVLLPKEAPRVARPERPRADTGHAFKAPKPARSGLEPASPSPKSGSE 1980
Db 1557 -----IDEDA--GRRHSWHDEDEAFDESP-----ELKYRETKSQE 1590
QY 1981 PRELPVPVGHATARTPAKNLAPHASPDPPAPPASAGDPHREKTSQKPFST-----QE 2035
Db 1591 SEELV--VTGGGLRRFKTIEL-----NSTIADKYSAESSQKTSLYFDEEPE 1636
QY 2036 LELRSL-----GYHGSYSYSGVEPVSPVSPSLTHDKGLPKHLEELDKSH 2081
Db 1637 LEMESLTDSPEDSRGEGSSLSHASTFT-----GTSPTSVSLSDED-----SDSSPSH 1685
QY 2082 LEGELRPKQPGVKLGAEAAHLPLRPSPSS-----SPLLTQAPGVKGHRVV 2133
Db 1686 KKGESKQR-----KARHRPHGLPLPTIENSSSEELREEEELKE-----QEKREI 1733
QY 2134 TLAQHTS-----EVITQDTRHHPOOLSAPLPAPLYSPFGACPVLDLRPP--SD 2182
Db 1734 EQQRKSSSKSKDKDELRAORRRERPK--TP-PSNLSPIEDAS-PTIELRQAAMEE 1788
QY 2183 LYLPPDPDHGAPRSGSHSGKRSPE-----PNKTSVLGGEGGIEPVSPPEGM 2231
Db 1789 LHRSSCSEYSPSTES-DPEGFEISPEKITEVQVKYKLPATAVSLYSPTDE--QSIMQKGS 1845
QY 2232 TEPGHSRSVYPLLYRDGEQ-----TEPSRMGSKSP-----GNT 2265
Db 1846 QKALKAAEYEMHMKHYKAFPAANERDEVEKEPLYGMLIEDYIESLIVETVNG 1905
QY 2266 SQPAPFSLKTESNAM-VKSKQKQKINKLNTNRNEPEYNSQPGTEIFNMPA-----2318
Db 1906 SVDGSLLTQREENGFMQKGRQKI--RLSEQIYEDPMQKITDLQKFEYELSHSVVP 1963
QY 2319 -----ITGTGLMTYSQAOVEASTNMGLEAIRKALMKGYDOWEESP 2362
Db 1964 QEDIVSSFTIIPESHEIVDLGTMTVTEERKLLDADAAYEELMKRQOM-----QLTP 2016
QY 2363 LSANAFNPL-----NASALPAAMPITAADGRSDHTLTSPGGGKAKVSG 2407
Db 2017 GSPTOAPTGEDMTDFRMPDASLTSSVLGASLTDSTSSATSLIP-----DVKITQ 2072
QY 2408 RPSRKAK-----SPAPGLASGRPPSVSVHSEGD 2439
Db 2073 HFSTEBIEEDVTDHREIQEIIAHESLILTYSEPSATSVPPSDTSLTSSVSVCTT 2132
QY 2440 NRRTPLTN-----RVVEDRPSAGSTFPFYNPLIM--RLQAGVMASTPPPP 2482
Db 2133 DSSSPITLDSITTVYTEPVDMITKPEDSEIISSTYFPGSIIDYPEEIASLDRTPAPD 2192

QY 2483 G-----LPAGSGPLAGPHHAWDEEPKPLCSCQVETLSDSE 2517
Db 2193 GRASADHIVISLSDMASSIIESVVPKPEGVA-----DTVSTDLLISEKDPVKKAK 2243

RESULT 90

ABR64281
ID ABR64281 standard; protein; 2468 AA.

XX ABR64281;

AC AC

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

XX DT 15-OCT-2003 (first entry)

Db 1862 GKTP-----GDFSYAQPEETTRSPDEDDYSEYKTRTSDVGYY 1905
QY 1943 -PRVARPERPRADTGHAFF-LAKPPARSGLEPASSPSKSGSEPRPLVPVPSGHATTARTPA 1999
Db 1906 YEKIERTTKSPDSGYSYETIGK-----TTKTPEDGDYSVEIIEKTT-----RTP 1951
QY 2000 KNLAPHASPDPPAPASADPHREKTQKPSIQLELRSLGYSYSPGEGVEPVSPV 2059
Db 1952 EGGYSYDISEKTTSP- EVSGYSYKTERSRLLDDI--SNGYDDSE--DQ----- 1997
QY 2060 SSPSLTHDKGLKHLSELDKSHLEGELPKQPGVKLGGEAAHLPHLRPLPSPQSSPL 2119
Db 1998 -----GHTLGDPSYSYETEK-----ITSPSEGSYS----- 2024
QY 2120 LOTAPGVKGQRVTLAQHISEVITQDYTRHHPPQQLSAPLAPLYSPFGASCPLDLRRP 2179
Db 2025 YETSTKTTTPDTSTVCYETAETKITR-----PQ-----ASTSYETSDLCYTAEKKS 2072
QY 2180 PSDL-----YLPDPDHGAPARGSPHSGGKSPBNKTSVLGGEDGIEPV 2226
Db 2073 PSEARQVDVLCVSSCEYKHPKTELSPFINPNLEWFASEPTESEKPLTQSGGAP-- 2130
QY 2227 PEGMTEPGHSNAVPLYRDGEQTEPSRMGSKSPCNTSQ--PPAFFSKLTESNSAMVK 2284
Db 2131 PPGGKQOG-----RQDETPTTSVSESAPSQTDSDVPPE-----TE----- 2167
QY 2285 SKQKEINKLNTHRNEPEYNISQPGTEIFNMPAITGTGLMTRVRSQ-----AVQE----- 2334
Db 2168 -ECPSITADANIDSEDE-----TIPTDKTVYKHMPPAPVQDRSPSP 2212
QY 2335 -HASTNM-----GLEAIIRKALMGKYDQWESPPPLSANAFNPLNASALPAAMPITAAD 2387
Db 2213 RHPDVMVDPEALAIQNGLKAL--KKDLKEKT-----KTKKPGTKTKSSPVKSD 2262
QY 2388 GRSDDLTLTSPGGGKAKVS---GRPSRRAKSPAPGLASGDRPPSVSVHSEGDGCMRTP 2444
Db 2263 GSKPLAASPKPAGLKESSDRVSRVASPKKGSVEKAAPPTTTPEVKAARGEEKETKN 2322
QY 2445 LTNRVWEDRPSAGSTPFPYVNLIMELQAGVMASPPPPGLPAGSGPLAGPHHA 2497
Db 2323 AANASAKSAKTATAGP-----GTTTKTKSSAVPGLPVYLDLCYIPNHS 2367

RESULT 91

AD62723
ID ADE62723 standard; protein; 2468 AA.

XX AC ADE62723;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein NP_005900, SEQ ID NO 8656.

XX KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002NO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX XX (GEO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

XX GENBANK; NP_005900.

PT New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the

XX polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2468 AA;

Query Match 3.1%; Score 410; DB 7; Length 2468;

Best Local Similarity 17.8%; Pred. No. 1e-13; Indels 892; Gaps 110; Matches 490; Conservative 360; Mismatches 1021;

QY 68 LLSEFQPGNERSQELHLPESHSLYPLGLG-----KSEMEFIESKPRLELLDPL 118
Db 174 LLSTTHPANKASLT-----FCPEEGDWKNSNLDHNLQDFINIKLSASILP---- 221
QY 119 RPSPLLATQAGSBDLTWDRSLTKLEPVSPSPHTDPELELVPP-----RLSKE-- 170
Db 222 -----EMGLSEFTEYLS-----ESVEVSPFDI-----LEPPTSGGFLKSKPCC 262
QY 171 -----ELIQNMDRVDEI-----T 184
Db 263 YIFPGRGDSALFVNGFNMLNGSERKSCFWKLIHLDVRDVSILLTHIGDNLPGINS 322
QY 185 MVEQOISLKKKQOQ---LEEBAK---PPE-----PEKPVSPPTIESKHSRLVQII 230
Db 323 MLORKIAELEEEQSGSTNSDWKNLISPDIGVFLVNPENLKNPEPNIKMKRSIEEAC 382
QY 231 YDEN--RKAAEAHRILEGLGPQVELPLYNQPSDTRQYHENTKINQAMKKLILYFKRN 288
Db 383 FTQLVNLKLSMKPEPLFRSVGNTIDPVILFQKMGVGL-EMVVLNPKSKEMQYFMQW 441
QY 289 HARKQWKQKFC---ORYDQLEALEKKVERIENPRRAK-----ESKVREY 332
Db 442 TGTNKDKAEFILPNGQEVLDPLISYITSVSSSLVWHPANPAEKIIRVLPFGNSTQYNILEG 501
QY 333 YEK-----QFP-EIRKQRELOERMOSRVGQSGSLGSMASAREHEV 372
Db 502 LEKLHLDLFLKQPLATQKDLTGQVTPVVKQTKLQORADSRSLKPAAPLPSKVRKES 561
QY 373 SEIIDGLSEQENLEKOMROLAVIPMLYDADQORIKFINNMGLMADPMKVIKDRQVMNMW 432

Db 1351 EVIEKPP-----AVP-----VSF 1363
Qy 1362 SYVEAQDYLRREAKLLKREGTPPPPPSRDLTEAVKTOALGPKLKP-----AHEG-L 1414
Db 1364 EFSDAKDN-----ERASVPMDEFPVD--SESPIEKVLPLRPPPLIGESAYESFL 1414
Qy 1415 VATVKEAGRSIHIPREELRHPTPELAPRPL-----KESGITQGTPLKYDTG 1462
Db 1415 SADDKASGRGA--ESPFEKSGKQSDQVSPVSEMTSTSLYQDKQSGKSTDPAPIKEDFG 1473
Qy 1463 ASTTSGKKHUVRSLLIGSPG-----RTFPVHP--LDVM-----AD 1495
Db 1474 QE----KKTDDVEAMSQPALALDERKLGDVSPQIDVQFSGFKEDTKMSISEGTVSDKS 1530
Qy 1496 ARALERACVEESLKSRPGTASGGSGIARGAPVIVPELCKPRQSPUTYEDHGAPFAGHLP 1555
Db 1531 ATPVDGVAEDTYSHMEGVASVSTASVATSS---FPEPTDDVSPSLHAEVSPSHSTEV 1587
Qy 1556 RGSPTVMREPTPLRQEGSLSSKASQDRKLTSTPREI-----AKSPHSTVPEHHHPISPY 1611
Db 1588 DLSLSVSVVQPTTTFQETENSPSKEECPRPMISPPDFSPKTAKS--RTPVQDHRSE-----1641
Qy 1612 EHLRLGVSGVDLYRSHIPLAFDPTSI PRGIPLDAAAAYLDRHLADNPYPHYLPYLIR 1671
Db 1642 -----QSSMSIEFGQESPEQSLAMDFS-----RQSPDHPVT-----1672
Qy 1672 GYPDTAALENROTILINDYITSQMHNTATAMAQRADMLRGLSPRESSLALVAAGPRGI 1731
Db 1673 GAGVLHITENGPTEV--DYSPSDM-----QDSSLHKIPPMEEP-----SITQDN-----1715
Qy 1732 IDLSQVPHLPVLVPTTGTATAMDRLAYLTAPOQPFSSRHS-----SSPLSGPGTHLPK 1787
Db 1716 -DLSEL-----ISVSQVEASPST-----SSAHTPSQIASPLQEDTLDVAP 1755
Qy 1788 PTTTS-----SSERRDRDRDREREKSLTSTTVEHAPIWRPTEQSGSGSGSG 1842
Db 1756 PRDMSLYASLTSEKVSLEGE-----KLSPKDISPLTPRESSPLYSPTFSDSTSAVKET 1811
Qy 1843 GGGGSSRRPASHAHQHPISPRITODALQQQPSVLHNTGMKIITAVEPSPKPTVLRSTS 1902
Db 1812 ATCHSSSSP-----PIDAASAEYGFASVLFTMQHHLALNRLDSTPGLEKOSG 1861
Qy 1903 TSPVPRPAATFPATHCPGLGTLGIVYPTLMPVLLPKEA-----1942
Db 1862 GKTP-----GDFSAYQKPEETTRSPDEEDYDSEYKTRTSDVGYY 1905
Qy 1943 -PRVARPERPRADTGAF--LAKPPARSGLEPASPSKGEPRPLVPVSGHATARTPA 1999
Db 1906 YEKIERTTKSPSDSGSYETIGK-----TTKTPEDGDYSYEIIIEKTT-----RTPE 1951
Qy 2000 KNLAHPHASPDPPAPASADPHREKTQSKPFSIQLELRLSLGYHSGSYSPGVEPVSFV 2059
Db 1952 EGGYSYDISEKTTSP--EVSGYSYKETERSRLLDDI-----SNGYDDSE---DG-----1997
Qy 2060 SSPSLTHDKGLPKHLBELDKSHLEGELRPKQPGVPLKGLGEAAHLPHLRPLPSPQSSSPL 2119
Db 1998 -----GHTLGDPSYSYETTEK-----ITSPSESGVS---2024
Qy 2120 LQAPGVKGHQVRVUUAQHISEVITQDYYRHHQQLUSAPLAPLYSPFGASCPVLDLRP 2179
Db 2025 YETSTKTRTPDTSTCYETAETKITRT-----PQ-----ASTSYETSCLCYTAEKKS 2072
Qy 2180 PSDL-----VLPDPDHGAPARGSPHSEGGKSPENKTSVLGGEDGIEPVS 2226
Db 2073 PBEARQDVLCVSSCEYKHPKTELSPFINGNPLEWFASERPTSEKEXPLTQSGAP--2130
Qy 2227 PPEGMTPEGHSRAVYPLLRYDGEQTEPRSMGSKSPGNTSQ--PPAFFSKLTESNSAMVK 2284
Db 2131 PPPGGKQOG-----RQCDTEPTTSVESAPSQDSDVPE-----TE-----2167
Qy 2285 SKQKEINKKLTHNRNPEYINISQPCETEFNMPAITGTGLMYRSO-----AVQE-----2334
Db 2168 -ECPSITADANIDSESE-----TIPTDKTVTKHMDPPPPAPVQDRSPSP 2212

Qy 2335 -HASTNM-----GLEAIIRKALMGKYDQWESPPLSANAFNPLNASASLPAAMPITAAD 2387
Db 2213 RHPDVSMVDPEALATEQNLKAL--KKDLKEKT-----KTKKPGTKTKSSSPVKKSD 2262
Qy 2388 GRSDHTLTSPPGGGKAKVS---GRPSSRKAKSPAPGLASGDRPPSVSVSHSGDCNRRTP 2444
Db 2263 GSKSPLAASPAPGLKESSDKVSRSVASPKKESVEKAAKPTTTTPEVKAARGEKOKETKN 2322
Qy 2445 LTNRWEDRPSAGSTPPFPNPLIMRLQAGVWASPPPCLPAGSGPLAGPHEA 2497
Db 2323 AANASASKSAKTATAGP-----GTTKTKSSAVPPGLPVYLDLCYIPNHS 2367

RESULT 94

ADB62715
ID ADE62715 standard; protein; 2468 AA.
XX AC ADE62715;
AC AC ADE62715;
DT 29-JAN-2004 (first entry)
XX DE Human Protein NP_005900, SEQ ID NO 8648.
XX KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GHEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
WPI; 2003-268312/26.
DR GENBANK; NP_005900.

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating


```
QY 1788 PTTTS-----SSRRDRDRDRDREREKSIILTSITTTVEHAPIWRPTEQSSGSSG 1842
Db 1756 PRMSLYASLTSEKVSQLEGE-----KLSPKSDISPLTPRESSPLYSPTSDSISAVKEXT 1811
QY 1843 GGGSSSRPASHAHQHSIPISPTQDALQORSVLHNTGKGIITAVPSPKPTVLRSTS 1902
Db 1812 ATCHSSSSP-----PIDAASAPYGFRAVLFDFTMHHALNRDLSTGLEKDSG 1861
QY 1903 TSSPVRPAATFPATHCPLGGTLGVVPLMEPVLLPKEA----- 1942
Db 1862 GKT-----GDSYAKQDEETRSRDEDDYSEYKTRTSDVGGY 1905
QY 1943 -PRVAPERPRADTGHAF--LAKPARSGLEPASPSKSGSEPLVPVPSGHATARTPA 1999
Db 1906 YEKIERTKSPSDSGSYETIGK-----TTKTPEDGDYSVEIIEKTI-----RTPE 1951
QY 2000 KNLAPHASDPDPAPASADPHREKTQSPFISQLELRSLGYHSGSYSPGVEVSPV 2059
Db 1952 EGYSYDISEKTTSPF-EVSGSYEKTERSRLLDDI---SNGYDDSE---DG----- 1997
QY 2060 SSPSLTHDKGLPKHLEELDKSHLEGELRPKQPGPVKLGGEAAHLPHLRPLPESOPSSPL 2119
Db 1998 -----GHTLGDPSYSYETTEK-----ITSPSESGVS----- 2024
QY 2120 LQAPGVKHQRVVTLLAQHISEVITQDTRHHPPQQLSAPLAPLYFPFGASCFLDLRRP 2179
Db 2025 YETSTKTRTPDTSTVCYETAETRT-----PQ-----ASTSYETSDLCYTAEEKS 2072
QY 2180 PSLD-----YLPDPHGPARGSPHSEGGKSPENKTSVLGGEDGIEPVS 2226
Db 2073 PSEARQVDLCVSSCEYKHPKTELSFINPFLWFASBEETSESEKPLTQSGAP-- 2130
QY 2227 PPGEMTEPHGSRSAVYLLRYRGEQTEPGRMGSKSPGNTSQ--PPAFFSKLTESNSAMVK 2284
Db 2131 PPGGKQGG-----RQDETFPTSVSEASPSQTSQDVPE-----TE----- 2167
QY 2285 SKQEKINKLNTNRNEPYNISQGTETFNWPAITGTGLMYRSQ-----AVQE----- 2334
Db 2168 -ECPSTADANIDSEDESE-----TIPTDKTVTKHMDPPAPVQDRSPSP 2212
QY 2335 -HASTNM-----GLEAIRKALMGKYQWESPPLSANAFNPLNASASLPAAMPTIAD 2387
Db 2213 RHPDVSMDVDEALAIENLIGKAL--KKDLKEKT-----KTKKPGTKTKSSPVKXSD 2262
QY 2388 GRSDHLTSPGGGKAKVS---GRPSRRKAKSPAPGLASGDRPPSVSVHSGDCNRRTP 2444
Db 2263 GRSKPLAASPAPKAGLESKVSRSVAPSKKESVEKAAKPTTTPVKAARGEKDEKTKN 2322
QY 2445 LTRVWEDRPASGTFPPYNNPLIMRLQAGVMASPPPGPLPAGSGPLAGPHHA 2497
Db 2323 AANASAKSAKTATAGP-----GTTTKTKSSAVPPGLPVVYLDLCYIPNHS 2367

RESULT 95
ABGI6636
ID ABGI6636 standard; protein; 2519 AA.
XX
AC ABGI6636;
DT
DT
XX 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #16627.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
PF
```

```
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX N-PSDB; AAS80823.
```

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 46995; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2519 AA;

Query Match 3.1%; Score 410; DB 4; Length 2519;
Best Local Similarity 17.8%; Pred. No. 1e-13;
Matches 490; Conservative 360; Mismatches 1021; Indels 882; Gaps 110;

```
QY 68 LLSEFQPCNERSQELHLPESHVLPGLG-----KSEMFIESKRPLELLPDL 118
Db 225 LLSTTHPANKASLT-----FCPEEGDWKNSLDRHNLQDFINIKLSASILP 272
QY 119 RPSPLLATGQAPAGSBDLTGKSLTGKLRPVSPSPPHDTPLELVPP-----RLSKE 170
Db 273 -----EMEGUSEFTEYLS-----ESVEVPSPFI-----LEPPTSGGFLKUSKPC 313
QY 171 -----ELIQNMDRVREI----- 184
Db 314 YIFPGRGDSALFAVNGFNMLINGSERKSCFWKLIRHLDRAVDSITLLTHIGDNLPGINS 373
QY 185 MVEQOISLKKKQOO---LEEEAAK---PPE-----PEKVPSPPIESKHSVLQII 230
Db 374 MLQRKIALEEEQSGSTNSDMKMLISPDGLVVVFLNVPENLKNPEFNIMKRSIERAC 433
QY 231 YDEN--RKKAAEAHRILEGLGPQVELPLYNQPSDTRQYHENIKINOAMRKLLYFKERN 288
Db 434 FTLQVNLKLSMKPPLFRSVGNTIDPVILFQMGVGKL-EMVVLNPKVSKSEMQYFMQOW 492
QY 289 HARKQWKQKFC-----QRYDQLMEALEKKVEIENPRRAK-----ESKVEY 332
Db 493 TGTNKDKAEFILPNGQEVDPISVLTSSVSLVWHPANPAEKIIRVLFPNGNSTQVNILEG 552
QY 333 YEK-----QFP-EIRKQRELQERMQSRVGQSGLSAARSSEHV 372
Db 553 LEKULHDLFKOPLATQKDLTGQVPTPVVKTKLQKQADRSRESLKPAKPLPFSKVRKES 612
```


Db	2264	RHPDVSWDP	PALAI	EQVLG	KAL--K	DLKEKT-----	KT	KPGTK	TSS	SPV	KKSD	2313																																														
Qy	2388	GRSDH	TLTSP	CGG	KAKV	S---GR	PSR	KAK	SPAG	LAS	GD	RP	PS	V	S	V	H	S	E	G	D	C	N	R	R	T	P	2444																														
Db	2314	GKS	PLA	ASP	KAG	PK	PAG	L	K	SSD	KV	S	V	AS	P	K	K	S	V	E	K	A	K	P	T	T	T	T	2373																													
Qy	2445	L	T	N	W	E	D	R	P	S	S	A	S	T	P	P	N	L	M	R	L	Q	A	G	M	A	S	P	P	P	P	G	L	P	A	G	S	G	P	L	A	G	P	H	A	2497												
Db	2374	A	A	N	A	S	A	S	K	A	T	A	G	P-----	G	T	T	K	T	K	S	A	V	P	P	G	L	F	V	I	D	L	C	I	P	N	H	S	2418																			
RESULT 96																																																										
ABO44398																																																										
ID	ABO44398 standard; protein; 2185 AA.																																																									
XX	ABO44398;																																																									
XX	26-SEP-2003 (first entry)																																																									
XX	Novel human protein kinase #18.																																																									
DE	Human; kinase; enzyme; cosmetic application; nutraceutical application.																																																									
KW	Homo sapiens.																																																									
OS	Homo sapiens.																																																									
XX	US6541252-B1.																																																									
PN	01-APR-2003.																																																									
PD	14-MAY-2001; 2001US-00854856.																																																									
XX	19-MAY-2000; 2000US-0206015P.																																																									
XX	(LEXI-) LEXICON GENETICS INC.																																																									
PA	Walke DW, Hilbun E, Donoho G, Turner CA;																																																									
XX	WPI; 2003-575927/54.																																																									
DR	N-PSDB; ACH03781.																																																									
DR	New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.																																																									
PT	Disclosure; Page; 11pp; English.																																																									
XX	The invention relates to a new isolated nucleic acid encoding a novel human protein kinase. The nucleic acid is useful in cosmetic and nutraceutical applications. The present sequence represents the amino acid sequence of a novel human protein kinase. Note: The sequence data for this patent did not from part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6541252B1																																																									
CC	Sequence 2185 AA;																																																									
XX	Query Match 3.1%; Score 408; DB 7; Length 2185;																																																									
XX	Best Local Similarity 19.6%; Pred.No. 1.1e-13;																																																									
XX	Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;																																																									
QY	147	FVSP	PP	PP	PH	T	D	P	E	L	V	P	R	L	S	K	E	L	I	Q	N	M	D	R	V	D	R	E	I	T	W	V	E	Q	I	S	K	L	K	K	Q	O	O	L	E	E	A	K	206									
Db	39	P	L	S	L	P	O	S	I	P	A	A	V	Q	S	A	P	E	P	H	R	E	T	V-----	T	A	T	S	O	V	A	Q	P	P	P	A	A	A	P	G	E	Q	A	V	A	89												
QY	207	P	P	E	P	P-----	E	K	P	V	S	P	P	P	I	S	K	H	R	S	L	V	Q	I	I	D	E	N	R	K	A	A	H	R	L	E	G	L	G	P	Q	V	E	L	P	L	P	257										
Db	90	G	P	A	S	T	V	P	S	T	S	K	D	R	P	V	S	P	S	L-----	V	G	S	K	E	E	P	P	P	A	121																											
QY	258	N	Q	P	S	D	T	R	O	Y	H	E	N	I	K	I	N	Q	A	M	R	K	L	I	L	Y	F	K	R	N	H	A	R	K	O	M	K	O	F	C	O	R	V	D	O	L	M	E	A	L	E	K	V	E	R	I	E	317
Db	122	R	S	G	S	G-----	G	G	S	A	K	E	P	O	E	R	S	O	O	D	I-----	E	E	L	E	T	K	A	V	G	M	S	155																									
QY	318	N	N	P	R	R	A	K	S	K	V-----	R	E	Y	E	R	K	O	F	P	E	I	R	K	O-----	E	L	Q	E	R	M	O	S	R	V	Q	O	R	G	S	G	L	S	M	A	A	R	S	E	369								

Db 1039 KRYHRSVRSRSHKTSRKLRLNVNKG-----RVVECOLETHNRKMTFKFD 1090
QY 1293 -DCRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPIPERHSPHLKEQHHI 1351
Db 1091 LOGDN-----PEETATWVNDPI-----LAIERESFVQVREII-----EKADEMLSEDSV 1138
QY 1352 RGSITGIPRVSVEAGEDYLRREAKLLKREGTPPPPPPPRDLTEAYKQALGPLKUKPAH 1411
Db 1139 EPEGDQGL--ESLQGHDDYGFSGSKLEGEFKQIPASSN-----PQIGIPT 1184
QY 1412 EGLVATVKEAGRS--IHEIPRELRITPELPLAPRLKESITQGTPLKYDTGASTTGSK 1469
Db 1185 SSLTQVHAGRRFVSPVESLRSKVPP-----SEIT-----DTVAASTAOS 1229
QY 1470 -----KHDVRSL-----IGSPGRFTFPVPHPLDVNMADARALE 1500
Db 1230 PGMNLSHASSLSLQAFSELRAQTEGNTAPPNFSHTGTFPVVP-----1278
QY 1501 RACYESLSKSPGTASSGGSIGARGAP-----VIVPELGKPRQSP--TYEDHGAPFA 1551
Db 1279 ---FLSSIAGVPTTAATAPVPATSPPNNDISTSVIQSEVTVPTERGIAGVATSTGVVTS 1335
QY 1552 GHLPRGSPVMTREPTRLQEGSLSSSKASQDRKLTSTPREIA---KSPHSTVPEHHPHI 1608
Db 1336 GGL-----PIPPVSESPLVSVSS-----ITIPAVVISITSPSLQVPTSTSEIV 1381
QY 1609 SPVEHLLRGVSGVDLYRSHRPIAFDPTSIPIRGIPDLDAAYLPRHLAPNPTYPHLYPPV 1668
Db 1382 -----VSSALYPS-----VTVSATSAGGSTATPCPK-----PPA 1413
QY 1669 LI-----RGYPDTAALENRQ-----TIINDVITSQQMH 1696
Db 1414 VVSQAAGSTTVGATILTSVTTSTFFSTASQLSLSSTSTPTLAETVVSASHLDKTS 1473
QY 1697 HNTATAMA-----QRADMLRGLSPR-----ESSIALNYAAGPRGIID 1733
Db 1474 HSTTGLAFSLASAPSSSSPGAGVSIVSQPGHPLVIPSIVASTPILPQAGAPTSTPL 1533
QY 1734 LSQVPHPLVLPPTGTPATAMDRLAYLPTAPOPPSSRRSSSPLSPGCP--THLTKTPTS 1792
Db 1534 LPQVPSIPLVQVANPAV--QOTLIHSQFP-----ALLNQPHTHCP-----1576
QY 1793 SSERDRDRDRDREREKSILTSITTVHEAPIWRPGTEQSGSGSGSGGSSSRPA 1852
Db 1577 -----EVDSDTOPKAPGIDDIKTLEE-----KLRSLSFSSHSSGA-----1611
QY 1853 SHSHAHQSPISPTODALQORPSVLHNTCMKI--ITAVEPSKPTVLRSTSSSPVRPA 1911
Db 1612 -----OHASVSLTS-----LVTESTVTGIPITTAVAPSK--LLTSTT-----S 1648
QY 1912 TFPPTHCPGLGGLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEP 1971
Db 1649 TCLPPTNLPLG-----TVALPVTVPVTPGQVSTPVTSTTSVGRP 1687
QY 1972 ASSPSKSGSPRLPVVPSVGHATARTPAKNLAPHASPPPPAPASDPHREKTS-KP 2030
Db 1688 GTAPSKPPLTKAPVLPVGTLPAGTLPSEQL-----PPFPGPSL-----TQSQP 1732
QY 2031 FSTQELERSLGLVHGSSYSPGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQ 2090
Db 1733 LEDLDAQLR-----RTLSPXITVTSV-----1755
QY 2091 PGVPKLGGEAAHPLHPLRIPE---SQPSSSPLLQATPV---KGHVVVTLAQHISEVI 2143
Db 1756 -GPNVNAAPTA--ITEAGTQKQVSVQKGVPLATSSGAGVFKMGFQVSVAA-----1806
QY 2144 TDYTRHHPQOLSAPLAPLYSPGASCPVLDLRRPPSDLYLPPP-----DHGAP 2193
Db 1807 --DGAQKEGKNKSEDAKSVHFESSTESSVLSSSPSTLVKPEPNGITIPGISSDVPS 1864
QY 2194 ARGSPHSEGGKRSPEPNKTS-----VLGGEDGIE-----PVSPPGEMTE 2233
Db 1865 AHKTTASEAKSDTQPTKVGRFOVTTTANKVGRFSVSKTEDKITDTKKGPPVSPFMDL 1924

RESULT 97

ABO44382

ID ABO44382 standard; protein; 2245 AA.

XX ABO44382;

AC ABO44382;

XX 26-SEP-2003 (first entry)

XX Novel human protein kinase #2.

XX Human; kinase; enzyme; cosmetic application; nutraceutical application.

XX Homo sapiens.

XX US6541252-B1.

XX 01-APR-2003.

XX 14-MAY-2001; 2001US-00854856.

XX 19-MAY-2000; 2000US-0206015P.

XX (LEXI-) LEXICON GENETICS INC.

XX Walke DW, Hilbun E, Donoho G, Turner CA;

XX WPI; 2003-575927/54.

XX N-PSDB; ACH03765.

XX New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.

XX Disclosure; Page; 11pp; English.

XX The invention relates to a new isolated nucleic acid encoding a novel human protein kinase. The nucleic acid is useful in cosmetic and nutraceutical applications. The present sequence represents the amino acid sequence of a novel human protein kinase. Note: The sequence data for this patent did not from part of the printed specification but was obtained in electronic format directly from USPTO at

XX seqdata.uspto.gov/sequence.html?DocID=6541252B1

XX Sequence 2245 AA;

XX Query Match 3.1%; Score 408; DB 7; Length 2245;

XX Best Local Similarity 19.6%; Pred. No. 1.1e-13;

XX Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;

QY 147 PVSPPSPHTDELELVPPRLSKBELIQNMDRVDRITVWEQOISKLKKKQOOLEBEAK 206

Db 99 PLSLPQSPISAPAAVPSQAPPEPHRETV-----TATATSQVAAQPPAAAAAGQAVA 149

QY 207 PREP-----EKVSPPIESKRSILVQIYDENRKAAEAHRLISGLGQVPELPLY 257

Db 150 GPAPSTVPSSTKDRVPSQSL-----VGSKEPPPA 181

Db 1867 --DGAQKEGKNKSEDAKSVHFESSTSESSVLSSTSSPESTLVKPEPNGITIPGSSDVPES 1924
Qy 2194 ARGSPHSEGGKSPENKTS-----VLGGEDGTE-----PVSPPEGMT 2233
Db 1925 AUKTTASEAKSDTGQTKVGRFOVTTANKVGRFVSUKTEKIDTKTKGVPVSPFMDL 1984
Qy 2234 PGHRSVAVLYLRDGEOTPEPSRMGSKSPGNTSQP--PAFFSKLTESNSAMVSKKQEIWK 2292
Db 1985 EQAVLPAVIPKKEKP--ELSEPSHLN---GPSSDPEAAFLSRDVEDDGGSGSPHSPHQLSSK 2039
Qy 2293 KLANTHRNPEPEYNISQPGTEIFNMPAITGTGLMYRSQAVQ-----EHAFTNMGLEA 2344
Db 2040 SL-----PSQNLSSQLSNSFNSYMSDNESDIEDKLELRRLRDKHLKEIQDLQS 2092
Qy 2345 IIRKALMGKYDOWEESPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGGGKAK 2404
Db 2093 RQKHETESLYTKLGKVP-----AVIIPPAAPLS-----GRRRRPKS-----KGS 2133
Qy 2405 VSGRPSSRKAKSP-AFGLASGRPPSV-----SSVHSEGD 2438
Db 2134 KSSRSSSLGNKSPQLSGNLGGQSAASVLPHPQOQLHPPGN 2172

RESULT 98
ID ABO44397
XX ABO44397;
AC ABO44397;
DT 26-SEP-2003 (first entry)
DE Novel human protein kinase #17.
KW Human; kinase; enzyme; cosmetic application; nutraceutical application.
OS Homo sapiens.
XX US6541252-B1.
XX 01-APR-2003.
XX 14-MAY-2001; 2001US-00854856.
XX 19-MAY-2000; 2000US-0206015P.
XX (LEXI-) LEXICON GENETICS INC.
PI Walke DW, Hilbun E, Donoho G, Turner CA;
XX WPI; 2003-575927/54.
DR N-PSDB; ACH03780.
XX New nucleic acid encoding novel human proteins, useful in cosmetic and
PT nutraceutical applications.
XX Disclosure; Page; 11pp; English.
PS The invention relates to a new isolated nucleic acid encoding a novel
CC human protein kinase. The nucleic acid is useful in cosmetic and
CC nutraceutical applications. The present sequence represents the amino
CC acid sequence of a novel human protein kinase. Note: The sequence data
CC for this patent did not from part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocId=6541252B1
XX Sequence 2322 AA;
SQ

Query Match 3.1%; Score 408; DB 7; Length 2322;
Best Local Similarity 19.6%; Pred.No. 1.2e-13;
Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;
Qy 147 PVSPSPPHPTDPELELVPPRLSKEELIQNMDRVDREITWVEQQISKLKKKQQLLEAAK 206

Db 39 PLSLPQSPISAPAAVPOSAPPPEHRETV-----TATATQVAAQPPAAAPGQAVA 89
Qy 207 PPEP-----EKVSPPIESKHSRLVQIYIDENRKKAAAHRLGLEGLGQVQLPLY 257
Db 90 GPAPSTVPSSTSKDRVSPSL-----VGSKEPPPPA 121
Qy 258 NOPSDTROYHENIKINQAMRKKLILYFKRNRHARKQWKQFCORYDQLEALEKKYERIE 317
Db 122 RSGSG-----GGSAPQPEERSQQDDI--BELETKAVGMS 155
Qy 318 NNPRRAKESKV--REYKQFPEIRKQ-----ELQERMOSRVQORSGLSMSARSE 369
Db 156 NDGRFLKFDIEIGRGSFKTVYKGLDTETTVEVAMCELODRK-----LTKSRORF 205
Qy 370 HVSSETIIDLGSQENLEKQMRQLAVIPMLYDADQORIK-----FINMNGML-ADPMKVYK 424
Db 206 KBAEMLKGL-QHPNIVR-----FYDSWESTVAGKKCIVLVTELMTSGTUKTYL 253
Qy 425 DR-----QVMNMWSEQEKETFREKFMQHPK-----NFGLIA 455
Db 254 KFKVMKIKVLSRWCQRILKGLQLHTRTPPIIHRDLKCDNIFITGTGTVKIGDLGL-- 311
Qy 456 SFLEKRTVAECVL-----YYLYTKQENY-----KSLVRSYRRRGKSQQOQQOQ 501
Db 312 ATLKGASPAKSVIGTPEFMAPEMYEKEYDESVDVYAFGMCLEMATSEYYPSECQNAQI 371
Qy 502 QOQQOQQOQP-----MPSRSQ-----BEKDEKEKEKEKEKEKEKEKEKEKEDLLK 547
Db 372 YRRVTSGVPASFDKVAIPEVKEIIEGCTRONKDERYSIKDLLNHAFFQETGVRVELAE 431
Qy 548 EKTDDTSGEDNDEKAVASKGRKTANSQGRKRITRSMANEANSE-EAITPQOQSAELAS 606
Db 432 E-----DDGEKIAIKLRLIEDIKKLGKYKDNEAIEFSPDLERDVPEDVAQ--- 478
Qy 607 MELNSSRWTEEMETAKGGLLEHGRNWSAIAARMVGSKTVSQCKNFYFNKKQKQNDL 666
Db 479 -EMVESGVYCEGDHKTMAKAIKDR-----VSLIK-----RKREORQLVR 516
Qy 667 QQHLKMEKERNARKKKA-----PAAASEEAAFPVVEDEMEASG 709
Db 517 EEQEKKQESLKKQOQVSSASQGIKQIPASCTIPTASTTSASVSTQVEPEPEA-- 574
Qy 710 VSGNEEMVVEAEALHASGNVPRGECSPATVN--NSSDTESIPSPHTEAAKDTQNGP 767
Db 575 -DOHQLOLQYQPSISVLSDGTVDSDGSSVFTESRVSSQTVSYGSGH-EQAHSTGTVP 632
Qy 768 KPATLGADGPPPG--PPT-----PPR 787
Db 633 HIPSTVQAQSPHGVYPPSSVAQSGOSQOPSSSLTGVSSSQPIQHPOQQOQGIQTAPPQ 692
Qy 788 RTSRAPIPT-PASEATGAPTPPPAPPSAPPVVPVPEKEEETAAAPVE--EGEOK 844
Db 693 QTVQVSLSTSTSSSEATTA-----QVSOQAPQVLPQVSAQKLPVSPVPTIQEPEQI 747
Qy 845 PPAASE--LAVDTGKAEPEVKSECTEEAEGBPAKGDAAEABAEATBEGALKAEKKGSGR 902
Db 748 PVATQSPVVPVHGAHFLPV-----GQPL 771
Qy 903 ATTAKSSGAPQSDSSATCSADEVBAEGDKNRLSPRSLTPTGDPANASPKPLD 962
Db 772 PTPPL-----LPQVPVSIPISTPHVSTAQTG-----PSSLFITMA-----AGITQPLL 814
Qy 963 LKQLKQRAAIP-----PIQVTKVHEPPREDAATPKAPPAPPPQNLPQESDAPQPG 1016
Db 815 TLAASATTAAPGVSTVPSQLPTLLQPVTO-----LPSQVHQLLQ-----AVQSMG 863
Qy 1017 SSPRGKRSRPAPPADKEAFAEAQKLPDPPCWTSGLPFPVPPPREVIKASHAPDPSAFS 1076
Db 864 I-----PANL-GQAAEVLSSGD--VLYQGFPPRLPPQY-----PGDSN 899
Qy 1077 YAPPGHPLPLGHDHTRPVL--PRPTINPPPLISSAKHPSV--LERQIGAISGMSV 1131

Db 900 IAPSSNVASVCIHST---VLXPPMPTEVLATPGYFFTVVQPYVESNLLVPMGVG---GGQV 954
QY 1132 QLHVPIYSEHAKAPVGPVTMGLPLPMPDKKLAPSGVKQEQLSRPGQAGPESLGV-----1186
Db 955 QVQPGGSLAQAPTSSQAV-----LESTQGV-----SQVAPAEFVAVAQQA 998
QY 1187 --PTAOEASVLRGTALGSGVPGGSIITKGIPTSTRVPSDSAITYRGSITHGTPADVLYKGTIT 1244
Db 999 TQPTTLASSV--DSASHDVASG-MSDG--NENVPSSG-----RHEGRTT 1038
QY 1245 RIIGEDSPSLRDLGRDRLSPKGVHIVYEGKGVHLSYEGGMSVTQCSKE-----1292
Db 1039 KRYHRSVRSRSHKTSRPLKILNVNKGD-----RVVECOLETHNRKQVTFKD 1090
QY 1293 -DGRSSGPPHETAAPKRTYDMMEGRVGRALISSASIEGLMGRAPRERHSPHLKQHHI 1351
Db 1091 LDGN---PEEATIMWNDFI-----LAIERESFVDQREII-----EKADEMLSEDVSV 1138
QY 1352 RGSITQCI PRSYVEAQEDYLRLREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLKPAH 1411
Db 1139 EPEGDQGL--ESLQKDDYGFSGSKLEGEFKQIPASSM-----PQIGIPT 1184
QY 1412 EGLVATVKEGRS--THEIPRELRHTPELAPRELKSGSITQGTPLKYDTGASTTSGK 1469
Db 1185 SSLTQVHVSAGRRFIVSPVESRLRESKVFP-----SEIT-----DTVAASTAQS 1229
QY 1470 -----KHDVRSLSL-----IGSPGRTEFPVPHLDVMADARALE 1500
Db 1230 PGNVLSHASSLSLQQAFLSELRAQTEGNTAPNFSTGTPFPVVP-----1278
QY 1501 RACYEESLKSRRPGTASSSGSARGAP-----VIVPELGRPROSPL--TYEDHGAPPA 1551
Db 1279 ---FLSSIAGVPTTAAATAPVATSPNDISTSVIQSEVTVPTBGIAGVATSGVVT 1335
QY 1552 GHLPRGSPVMTREPTRLQEGSLSSSKASQDRKLTSTPREIA---KSPHSTVPEHHPHPI 1608
Db 1336 GGL-----PIPVSESPVLSSVSS-----ITIPAVSISTTSPSLQVPTSTSEIV 1381
QY 1609 SPYEHLLRGVSDVLYRSHIPLAFDPTSIPIRGIPDLAAAAYLPRHLAPNTPVPHLYPY 1668
Db 1382 -----VSSTALPS-----VTVSATASAGGSTATPGPK-----PPA 1413
QY 1669 LI-----RGYPTTAALNRQ-----TIINDYITSQOMH 1696
Db 1414 VVSQQAAGSTTVGATLTSVSTTSPSTASQLSISSTSTPTLAETVVSASHLSDKTS 1473
QY 1697 HNTATAMA-----QRADMLAGLSR-----ESSALNAYAAGRGDI 1733
Db 1474 HSTTGLAFSLAPSSSSSPGAGVSSYISQPGGLHPLVIPSVIAGTPIPLFQAAGTSTPL 1533
QY 1734 LSQVPHLPVLVPTTGTATAMDRLAYLPTAPQFSSRHSPLSPGGP--THLTKTPTTS 1792
Db 1534 LQVPSIPLVQPVANVAV---QQTLIHSQGP-----ALLFNQPHTHCP-----1576
QY 1793 SSERDRDRDRDRDREREKSLTSTTVEHAPIWRPGTEQSGSGSGSGSGSGSSRPA 1852
Db 1577 -----EVDSDTQPKAPGIDDKITLEE-----KLRSLFSEHSSGA-----1611
QY 1853 SHSHAHOSPIPRQTQDALQORPSVLHNTGMKI--ITAVEPSKPTVLRSTSTSSVPRPA 1911
Db 1612 -----QHASVLETS-----LVIESVTGIPPTAVAPSK--LLTSTT-----S 1648
QY 1912 TFPPTHCPGLGLDGVYPTLMEPVLLPKAEPRVARPERPRADTGHAFIAKPPARGLEP 1971
Db 1649 TCLPPTNLPLG-----TVALPVTVPVTPQVSTFVSTTSGVKP 1687
QY 1972 ASSPSKGSFPLVPPVSGHATARTPAKVLAPHASDPDPAPPASADPHREKTS--KP 2030
Db 1688 GTAPSKPPTUKAPVLFPVGTLPAGTLPSEQL-----PPPPGFSL-----TQSQOP 1732
QY 2031 FSIQIELELASLGHSGSSVSPGVEPVSPVSPSLTHDKGLPKHLELDKSHLEGELRPKQ 2090
Db 1733 LEDLDAQLR-----RTLSPXITVTSAV-----1755

QY 2091 PGVVLGGEEAHLPHLRPLPE---SOPSSSPLLOTAPGV---KGQRVVTUQAHISEVI 2143
Db 1756 -GPFVSNAAFTA--ITRAGTQPKGVSVQKRGFPVLATSSGAGVFMGRFQVSVAA-----1806
QY 2144 TQDYTRHHPOQLSAPLAPLYSFPGASCPVLDRRPPSDLYLPPP-----DHGAP 2193
Db 1807 --DGAQKGGKNSDAKSVHFESSTSSSVLSSSPFSTLVKPEPNGITIPGISSDVPES 1864
QY 2194 ARGSPHSEGGKRSPEPNKTS-----VLGGEGDIE-----PVSPPEGMTE 2233
Db 1865 AHKTITASEAKSDTQGTQVGRFQVTTTANKVGRFSVSKTEDKITDTKKEGVPASPPFMDL 1924
QY 2234 PGHSRSAYVPLLYRDGEOTEPSRMGSKSPGNTSQP--PAFFSKLTSNAMYKSKQKQINK 2292
Db 1925 EQAVLPAPVIPKKEKP--ELUSEPSHLN-----GPSSDPEAAFLSRDVEDDGSSPHSPHLSK 1979
QY 2293 KLNTNRNEPEYNIQSOPGTEIFNMPAITGTGLMTYRQAVO-----EHASTNMGLEA 2344
Db 1980 SL-----PSQNLSSQLSNSFNSSVMSDSNEDIEDLKLRLRLDKHLKEIQDLOS 2032
QY 2345 IIRKALMKYQDWEESPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTPSGGGGKAK 2404
Db 2033 RQKHEIESLYTKLGVPP-----AVIIPPAAPLS---GRRRPTKS-----RGS 2073
QY 2405 VSGRPSSRKAKSP--APGLASGDRPPSV---SSVHSEGD 2438
Db 2074 KSSRSSSLGNKSPQLSGNLGQSAASVLHPQOITLHPPGN 2112

RESULT 99
ABO44381
ID ABO44381 standard; protein; 2382 AA.
XX ABO44381;
XX DT 26-SEP-2003 (first entry)
XX Novel human protein kinase #1.
XX Human; kinase; enzyme; cosmetic application; nutraceutical application.
XX Homo sapiens.
XX US6541252-B1.
XX PD 01-APR-2003.
XX PF 14-MAY-2001; 2001US-00854856.
XX PR 19-MAY-2000; 2000US-0206015P.
XX PA (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Hilbun E, Donoho G, Turner CA;
PI WPI; 2003-575927/54.
DR N-PSDB; ACH03764.
XX New nucleic acid encoding novel human proteins, useful in cosmetic and
PT nutraceutical applications.
XX Disclosure; Page; 11pp; English.
XX The invention relates to a new isolated nucleic acid encoding a novel
CC human protein kinase. The nucleic acid is useful in cosmetic and
CC nutraceutical applications. The present sequence represents the amino
CC acid sequence of a novel human protein kinase. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6541252B1
XX Sequence 2382 AA;

Query Match	3.1%;	Score 408;	DB 7;	Length 2382;	
Best Local Similarity	19.6%;	Pred. No. 1.2e-13;			
Matches	513;	Conservative	291;	Mismatches	943;
				Indels	872;
				Gaps	114;
QY	147	PVSPPPTHTDELELVPPRLSKELIQNDVRDREITVVEQIQISKKKQOOLEEAAK	206		
DB	99	PLSLPQPSIPAAVQSPAPPEHRETV-----TATATSOVAQPPAAAPGEQAVA	149		
QY	207	PREP-----EKVSPPEIESKHRSVLVQIYDENRKAABAAHILEGLGQVPELPLY	257		
DB	150	GRAPSTVPSSTKDRVPSQSL-----VSKKEPPPA	181		
QY	258	NOPSOTROYHENIKINOAMRKKLILYFKRNHARKQWKQFCORYDQLEALKKVERIE	317		
DB	182	RSGSG-----GGSAKEPQEBERSQOQDDI--EELETKAVGMS	215		
QY	318	NMPRRAKESKV-REYVEKQFPEIRKQ-----ELQRMQSRVQORSGLSMSAARSE	369		
DB	216	NDGRFLKFDIEGRGSKFTYKGLDTEVVEAWCLODRK-----LTKSERQRF	265		
QY	370	HEVSEITDGLSEENLEKQMLAVIPMLYDADQORIK-----FINMGLM-ADPMKVYK	424		
DB	266	KBAEAMKGL-QHPNIVR-----FYDSWESTVKGKICIVLTELMTSGTLKTYL	313		
QY	425	DR-----QVMNMWSQEKETFREKFMHPK-----NFGLIA	455		
DB	314	KRPVKMKIKVLRSMCHQILKGLQLHLRTPPIIHRDLKCDNIFITGPTGSVKIGDLGL--	371		
QY	456	SFLERTVAECVL-----YYLYTKQENY-----KSLVRSVRRRGKSGQOQOQOQ	501		
DB	372	ATLKRAFSFAKSVIGTPEFMAPEYEBKYDESVDVYAFGCMLEMATSEYPISECQNAQI	431		
QY	502	QOQOQOQOQO-----MRSQQ-----BEKDEKEKEKEBEKEBEVENDKEDLK	547		
DB	432	YRVTVSGVPASFDKVAIBEVEKIIIEGCTRONKDERISKOLLNHAFFQEBETGVRVELAE	491		
QY	548	EKTDDTSGEDNDEKAVASGRKNTANSQGRKRITRSMANEANSE-EAITPQCSAELAS	606		
DB	492	E-----DDGEKIAIKLWLRIEDIKLKGKYNDEAIEFSFLDRDVEDVAQ----	538		
QY	607	MELNESRWTEEBEMETAKGLLEHGRNWSAIAARMVGSKTVSQCKNFYFNKGRQNDLIL	666		
DB	539	-EMVESGYCEGDHKTMAKIDR-----VSLIK-----RKREQRQLVR	576		
QY	667	QOHLKWEKERNARRKKKA-----PAAASEEAFFPPVVEDEMEASG	709		
DB	577	ESQEKKKQBESSLKQOQVQSSASQTGIKQLPSASTGIPASTTSASVSTQVEPEPEA--	634		
QY	710	VSGNEEMVEEAEALHASGNEVPRGCSGPATVN--NSSDTEIPSPHTEAAKDTQONGP	767		
DB	635	-DOHQQLYQOQPSISVLSDGTVDGSGSVFTESRVSSQOQTVSGSQH-EQAHSTGVPG	692		
QY	768	KPATLGADGPPG--PPT-----PPR	787		
DB	693	HIPSTVQAOQPHGVPPSPSAQOQSGQSGSSSLTGVSSSQPIQHPQOQOQIQOATAPQ	752		
QY	788	RISRARIET-PASEATGAPTPPAPSPSPAPPPVVPKKEKEEATAAPVE--EGEOK	844		
DB	753	QTVQYLSQTSSTSEATTA-----QVVSQAPQVLPQVSAGKQLPVSPVPTIQGEPOI	807		
QY	845	PPAAEE--LAVDTGKAEEPVKSECTEABEGPAKDAEAAEATAEGALKAEKKGSGR	902		
DB	808	PVATQSPVPHSGAFLPV-----GQPL	831		
QY	903	ATPAKSSGAPQSDSDSATCADVEBAEGDKNRLSPRSLLTPTGDPANASPOKPLD	962		
DB	832	PTPL-----LPQVPSQIPISTPHVSTAQTG-----FSSLIPITMA-----AGITQPLL	874		
QY	963	LKQKORAAAI-----PQVTKVHEPPREDAPTKPAAPPAPPQNLQPEDSAPQCG	1016		
DB	875	TLASSATTAAIPGVSTWVPSQLPTLIQPVTO-----LPQVHPQLQOP-----AVQSMG	923		

QY	1017	SSPROKSRSPAPPADKEAPAAEAQKLPDPDPWCTSGLPFPFPPPREVIKASPHAPDSARS	1076		
DB	924	I-----PANK-GQAAEVPLSSGD--VLYQGFPLRPQY-----EGSDN	959		
QY	1077	YAPPCHPLPLGLHDTARPVL--PRPTISNPPLISSAKHPSV--LERQIGAISQMSV	1131		
DB	960	IAPSSNVASVCIHST---VLXPPMPTEVLATPGYPTVQPVVSVESNLLVPMGV--GGQV	1014		
QY	1132	QJHVYSEHAKAPGVPTVMGLPLPMDPKLAPFSGQKQBLSPRGQAGPPESLGV----	1186		
DB	1015	QVSPGGSLAQAPTSSQAV-----LESTQGV-----SQVAPAEPAVAQPOA	1058		
QY	1187	--PTQOEAIVLRGTALGSPGGSIKGPSTRTVPDSASITYRGSITHGTPADVLVYGTIT	1244		
DB	1059	TQPTTLASSV--DSAHSDVASG-MSDG--NENVPSSSG-----RHEGRTT	1098		
QY	1245	RIIGEDSPSLDRGREDSLPKGHVIEGKGVLSYEGGMSVTOCSKE-----	1292		
DB	1099	KHYRKSVRSRHREKTSRPLRIILNVSNKGD-----RVVECOLETHNRKMWTFKPD	1150		
QY	1293	-DCRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIPPERHSPHHLKEQHII	1351		
DB	1151	LOGDN-----PEIATIMVNDFI-----LAIERESFVQVREII---EKADMLSEDVSV	1198		
QY	1352	RGSITQGIPIRSYVBAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLPKH	1411		
DB	1199	EPEGDQGL--ESLQKDDYFGSGKLEGEFKQPIPASSM-----PQIGIPT	1244		
QY	1412	EGLVATVKEAGS--IHEIPRELHRTPELPLAPRLKEGSIQTGTPLYKVDGASTTGSK	1469		
DB	1245	SSLTQVTHSAGRRFVSPVPSRLRESKVFP-----SEIT-----DTVAASATOS	1289		
QY	1470	-----KHDVRSI-----IGSPGRTPFPVPHPLDVNMADARALE	1500		
DB	1290	PGMNLSSHASSLSLQAFSELRAQTEGTPAPPNFSTGTPFPVPP-----	1338		
QY	1501	RACYEESLKRPGTASSSGSIARGAP-----VIVPLGKPRQPL--TYEDHGAPFA	1551		
DB	1339	---FLSSIAGVTTAAATAPVATSPNDISTSVIQSEVTVPTTEGIAGVATSTGVVTS	1395		
QY	1552	GHLPRGSPVMEPTPRLOEGSLSSSKASQDRKLTSTPREIA---KSPHSTVPEHHPHI	1608		
DB	1396	GGL-----PIPPVSESPVLSVSS--ITIPAVSISTSPSLQVSTSTSEIV	1441		
QY	1609	SPVEHLLRGVGVLYRSHIPLAFDPTSPRGIPLDAAAAYLPRHLAPNPTYPHLYPPY	1668		
DB	1442	-----VSTALYPS-----VTVSATSAGSGSTATPGPK-----PPA	1473		
QY	1669	LI-----RGYPDTAALENRQ-----TIINDYITSQOMH	1696		
DB	1474	VVSQAAGSTTVGATLTSVSTTSPSTASQSLISLSSSTSTPTLAETVVVSAHSLDKTS	1533		
QY	1697	HNTATAMA-----QADMLRGLSPR-----ESSIALNYAAGRGID	1733		
DB	1534	HSSTTGLAPSLAPSSSSSPGAVSISOPGGLHPLVIPSVIASPTILLPQAAGTSTPL	1593		
QY	1734	LSQVPHLPVLVPTGTPATAMDRLAYLPTAPQFSSRSHSSPLSPGPG--THLTKTPTS	1792		
DB	1594	LPQVPSIPFLQPVANPAV---QOTLIHSQFP-----ALLNQPHTHCP-----	1636		
QY	1793	SSERDRDRDRDREREKSIILTSTTVVEHAPIWRPGTEQSSGSGSGSGGSSSRPA	1852		
DB	1637	-----EVDSDTPKAPGIDDIKTLEE-----KLRSFSEHSSGA-----	1671		
QY	1853	SHSHAHOSPISPRTQDALQQRPSVLHNTGMKI-ITAVEPSKPTVLRSTSTSPVPRAA	1911		
DB	1672	-----QHASVSLETS-----LVISTVTPTGIPITAVAPSK--LLTSTT-----S	1708		
QY	1912	TPPPATHCPGLGTLDGVPYPTLMEPVLLPKAPRVARPERPRADTGHAFKAPPARSGLEP	1971		
DB	1709	TCLPPTNPLFG-----TVALPVTVTPGVSTVSTTSVGVKP	1747		
QY	1972	ASSPSKSGSEPRFLVPPVPSGHATITARTPAKNLAPHASPPPPASADPHREKTOS-KP	2030		

Db 1748 GTAPSKPLTKAPVLPVGTLPAGTLLPSEQL-----PPFPCLP-----TQSQOP 1792
Qy 2031 FSTQELRLSLGVHGGSSYSPGVEPVSPVSSPSLTHDKGKLPKHELELDKSHLGBELRPQ 2090
Db 1793 LEDLDAQR-----RTLSPEXITVTTSV-----1815
Qy 2091 PGPKVLGGEAAHLPLRLPE---SOPSSPPLLQATPGV---KGHVVVTLAQHISEVI 2143
Db 1816 -GPVMAAPTA-ITEAGTQPKGVSVQKGPVLTSSGAGVFKWGFQVSVAA-----1866
Qy 2144 TDYTRHHPQOLGAPLPAPLYSPFGASCPVLDLRRPPSDLYLPPP-----DHGAP 2193
Db 1867 --DGAQKEGKNKSEDAKSVHFESSTSESSVLSSSSPESTLVKPEPNGITIPGISDVPS 1924
Qy 2194 ARGSPHSEGGKRPENKTS-----VLGGGEDGIE-----PVSPPGMTE 2233
Db 1925 AHKTTASEAKSDGTQPKVGRFOVTTANKVGRFSVSKTEDKITDKGPGVASPPFMDL 1984
Qy 2234 PHSRSVAVYLLYRDGEQTEPSSRMGSKSPGNTSQP--PAFFSKLTESNMAVSKKQEI 2292
Db 1985 EQAVLPAVIPKKEP-ELSEPHLN---GPSSDPEAFLSRDVGSGSPHPSHOLSK 2039
Qy 2293 KLTHNRNEPEYNISQGTETFMNPAITGTLMTYRSQAVQ-----EHASTNMGLEA 2344
Db 2040 SL-----PSQNLSQLNSFNSSYMSDSEDIEDLKLRLRLDKHLKEIQDLQS 2092
Qy 2345 IIRKALMGKYDOWEESPPLSANAFNPLNASASLPAAMPITADGRSDHTLTSPGGGKAK 2404
Db 2093 RQHEIESLYTLKGVPP-----AVIIPPAAPLUS---GRRRRPTKS-----KGS 2133
Qy 2405 VSGRPSRKAKSP-AFGLASGDRPPSV---GSVHSEGD 2438
Db 2134 KSSRSSSLGNKSPQLSGNLSSGQSAASVLPHPQQLTHPPGN 2172

RESULT 100
ID AAM41059
XX ID AAM41059 standard; protein; 1565 AA.
AC AAM41059;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5990.
XX
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60215.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
PS Example 2; SEQ ID NO 5990; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM39642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1565 AA;
Query Match 3.1%; Score 406.5; DB 4; Length 1565;
Best Local Similarity 20.8%; Pred. No. 8.7e-14;
Matches 368; Conservative 188; Mismatches 612; Indels 597; Gaps 82;
Qy 355 GORGSLMSAARSEHVSEIIDG---LSQENLEKQMKQLAVIPPLMDADQOIRKI 410
Db 10 GARGSGWG-SRSQAPYGTGLGAVSGGEQVLLHEEAGDSGFVLSRLGPSLRDKLEMEELM 68
Qy 411 NMNGLMADPMKVYKDRVMNMWSEKEKTFREKFMQHPKFNGLIASFLERKTVACVLYY 470
Db 69 LQETLLTGTQSYMDASLISLIEDFGSLGEVEMSLPDPSPFPLETS-----119
Qy 471 YLTKNENYKSLVRRSRRRGKSSQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 510
Db 120 --SPKLPSMRP--PRSRFRWQSPPPQQRSDGEEVEVAFSGQILAGELNCVSSIPDF 175
Qy 511 PMPRSSQKE 545
Db 176 PMHLACPEED-----KATAAEMAVPAAGDSSISLSSELVRAHPYCLPNLTHLASLEDE 230
Qy 546 LKEKTDD-----TSGED-----557
Db 231 LQEQPDDLTLPQGCVVLEIVQQAATAGDDLEIPVVVQVSPGPRPVLDDSLTSSALQL 290
Qy 558 -----NDEKEAVASKGRKTANSQGRKGRITRSMANEANSEAI-----596
Db 291 LMPTLESETEAAVPKVTLCSEKGL-----SLNSEKLDLSACLLKPREVPEPV 339
Qy 597 -----TPQOSAEIASMELNNESSRWTEEMETA-KGGLLEHGRNWSAIAWMVGSQK 650
Db 340 PKFQNPANAAAGSORARKGRKKKKEQPAACVEGYARRLRSSRSGQSTVGTEVTSQ-- 397
Qy 651 NFYFNKKRQNLDRILQOHKLMKERNARKKKKAPAAASEEAAFPVVVEDEMEASGV 710
Db 398 -----VDNLQKQPELOKESGPGLOGKPKRAWARAA-----ALENSSP 438
Qy 711 SGNNEEMVEAEALHASNEVPRGECSPATV--NNSSDT--ESIPSPHTEAAKDTGQNGPK 768
Db 439 KNLE-----RSAGQSSPAKE--GPLDLVPKLADTIQTNPIPTHLSLVDSQAQSPM 486
Qy 769 PPATLGADGPPGP-----PTPP-----RRTSRAPTEPTPASEATGATPPPPAPPSA 817
Db 487 PVDSEADPTAVGPVLGAPVPDPLGLVDLASTSSSELVEPLPAEPVL---INPVLADSAV 543

This Page Blank (uspto)